

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2003, 12:55:48 ; Search time 41 Seconds
(without alignments)
998.816 Million cell updates/sec

Title: US-09-813-453B-2

Perfect score: 1335

Sequence: 1 LLLVIDGNTWTVLGVYHDG.....PFLTKGLELIYERNRVGSV 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq 19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1335	100.0	258	AAU01243	B. subtilis novel
2	1335	100.0	258	AAU91149	Bacillus subtilis
3	1114	83.4	233	AAU91163.	Pantothenate kinas
4	1046	78.4	258	AAU91172	Pantothenate kinas
5	1034	77.5	262	AAU91170	Pantothenate kinas
6	987	73.9	254	AAU91171	Pantothenate kinas
7	852	63.8	259	AB47661	Listeria monocytog
8	794.5	59.5	256	AAU91175	Pantothenate kinas
9	756	56.6	255	AAU91154	Geobacter sulfurre

10	655	49.1	265	23	AAU91151	Streptomyces coeli
11	646.5	48.4	250	23	AAU91150	Clostridium acetob
12	577	43.2	260	23	AAU91173	Pantothenate kinas
13	539.5	40.4	256	23	ABP65945	Bifidobacterium lo
14	537	40.2	258	23	AAU91153	Rhodobacter capsul
15	515	38.6	219	23	AAU91176	Pantothenate kinas
16	494.5	37.0	272	23	AAU91152	Mycobacterium tube
17	493.5	37.0	272	23	AAG81225	Mycobacterium tube
18	428	32.1	262	23	AAU91155	Deinococcus radiop
19	405.5	30.4	246	23	AAU91156	Thermotoga maritim
20	338.5	25.4	212	23	AAU91177	Pantothenate kinas
21	327.5	24.5	273	23	AAU91157	Treponema pallidum
22	319	23.9	257	23	AAU91174	Pantothenate kinas
23	271.5	20.3	262	23	AAU91158	Borrelia burgdorfe
24	207	15.5	244	23	AAU91168	Pantothenate kinas
25	203	15.2	241	23	AAU91179	Pantothenate kinas
26	163	12.2	249	23	AAU91182	Pantothenate kinas
27	163	12.2	257	23	AAU91180	Synochocystis pant
28	154.5	11.6	229	23	AAU91159	Aquifex aeolicus p
29	154	11.5	249	23	AAU91178	Pantothenate kinas
30	151	11.3	455	20	AAV38617	Neisseria gonorrhoe
31	151	11.3	455	21	AAV74908	Neisseria gonorrhoe
32	151	11.3	460	23	AAU91167	Pantothenate kinas
33	151	11.3	592	20	AAV38618	Neisseria gonorrhoe
34	151	11.3	592	21	AAV74911	Neisseria gonorrhoe
35	151	11.3	592	24	ABP77010	N. gonorrhoeae ami
36	150	11.2	389	21	AAV74909	Neisseria meningit
37	150	11.2	455	21	AAV74910	Neisseria meningit
38	150	11.2	592	20	AAV38615	Neisseria meningit
39	150	11.2	592	20	AAV38616	Neisseria meningit
40	150	11.2	592	21	AAV74912	Neisseria meningit
41	150	11.2	592	21	AAV74913	Neisseria meningit
42	150	11.2	592	23	AAU91166	Pantothenate kinas
43	150	11.2	592	23	AAU91169	Pantothenate kinas
44	150	11.2	592	24	ABU06036	N. meningitidis va
45	134.5	10.1	242	23	AAU91180	Pantothenate kinas

ALIGNMENTS

RESULT 1

AAU01243

ID AAU01243 standard; Protein; 258 AA.

XX AC AAU01243;

XX DT 18-JUL-2001 (first entry)

XX DE B. subtilis novel pantothenate kinase encoded by the gene coaX.

XX DE Pantothenate kinase; coaX; pantothenate biosynthesis; vitamin B5;

XX DE nutritional supplement; panto-compound; pantoate.

XX OS Bacillus subtilis.

XX PN WO200121772-A2.

XX PD 29-MAR-2001.

XX PR 21-SEP-2000; 2000WO-US25993.

XX PR 21-SEP-1999; 99US-0400494.

XX PR 07-JUN-2000; 2000US-0210072.

XX PR 28-JUL-2000; 2000US-0221836.

XX PR 24-AUG-2000; 2000US-0227860.

XX PA (OMNI-) OMNIGENE BIOPRODUCTS.

XX PI Yocum RR, Patterson TA, Hermann T, Pero JG;

XX XX WPI; 2001-218644/22.

XX DR N-PSDB; AAS00984.

XX DR

(OMNI-) OMNIGENE BIOPRODUCTS INC.

Yocum RR, Patterson TA;
WPI; 2002-269358/31.
N-PSDB; ABK54169.
Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -
Disclosure; Page 81-82; 128pp; English.
The invention describes assays for identifying a (potential) antibiotic kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Sequence 233 AA;

Query Match 83.4%; Score 1114; DB 23; Length 233;
Best Local Similarity 99.5%; Pred. No. 5.7e-117;
Matches 213; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 LLLVIDGNTNTVLGVHDGKLEYHWRIETSRHKTDEFGMLRSLFDHSGLMFEQIDGI 60
Db 1 MLLVIDGNTNTVLGVHDGKLEYHWRIETSRHKTDEFGMLRSLFDHSGLMFEQIDGI 60
Qy 61 IISVVPPIPFALERMCTKYFHIETSPQIVGPGMKTGLNPKYDNPKEVGADRIWNAVAATHL 120
Db 61 IISVVPPIPFALERMCTKYFHIETSPQIVGPGMKTGLNPKYDNPKEVGADRIWNAVAATHL 120
Qy 121 YGNPLIVVDFTATTTCYIDENKQYMGGAIAFGITISTEALYSRAAKLPRIETRPDNI 180
Db 121 YGNPLIVVDFTATTTCYIDENKQYMGGAIAFGITISTEALYSRAAKLPRIETRPDNI 180
Qy 181 GKNTVSAMQSGILFGYGVQVEGIVKRMKWQAKQD 214
Db 181 GKNTVSAMQSGILFGYGVQVEGIVKRMKWQAKQD 214

RESULT 4

AAU91172
ID AAU91172 standard; Protein; 258 AA.
AC AAU91172;
XX
XX
XX 05-JUN-2002 (first entry)
XX
XX Pantothenate kinase (Coax) #10.
XX
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Bacillus stearothermophilus.
XX
XX WO200216601-A2.
XX
XX 28-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-US26531.
XX
XX 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
PA

Yocum RR, Patterson TA;

WPI; 2002-269358/31.
N-PSDB; ABK54193.
Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -
Claim 10; Page 101-102; 128pp; English.
The invention describes assays for identifying a (potential) antibiotic kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Sequence 258 AA;

Query Match 78.4%; Score 1046; DB 23; Length 258;
Best Local Similarity 78.3%; Pred. No. 3.1e-109;
Matches 198; Conservative 32; Mismatches 23; Indels 0; Gaps 0;
Qy 1 LLLVIDGNTNTVLGVHDGKLEYHWRIETSRHKTDEFGMLRSLFDHSGLMFEQIDGI 60
Db 1 MIFVLDGNTNTVLGVHDGKLEYHWRIETSRHKTDEFGMLRSLFDHSGLMFEQIDGI 60
Qy 61 IISVVPPIPFALERMCTKYFHIETSPQIVGPGMKTGLNPKYDNPKEVGADRIWNAVAATHL 120
Db 61 IISVVPPIPFALERMCTKYFHIETSPQIVGPGMKTGLNPKYDNPKEVGADRIWNAVAATHL 120
Qy 121 YGNPLIVVDFTATTTCYIDENKQYMGGAIAFGITISTEALYSRAAKLPRIETRPDNI 180
Db 121 YGNPLIVVDFTATTTCYIDENKQYMGGAIAFGITISTEALYSRAAKLPRIETRPDNI 180
Qy 181 GKNTVSAMQSGILFGYGVQVEGIVKRMKWQAKQD 240
Db 181 GKNTVSAMQSGILFGYGVQVEGIVKRMKWQAKQD 240
Qy 241 LTLGLLELIYERN 253
Db 241 LTLGLLELIYERN 253

RESULT 5

AAU91170
ID AAU91170 standard; Protein; 262 AA.
AC AAU91170;
XX
XX 05-JUN-2002 (first entry)
XX
XX Pantothenate kinase (Coax) #8.
XX
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Bacillus anthracis.
XX
XX WO200216601-A2.
XX
XX 28-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-US26531.
XX
XX 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX


```

11-APR-2000; 2000FR-0004629.
(INSP ) INST PASTEUR.
Buchrieser C, Frangeul L, Couve E, Ruaniock C, Fsihi H, Dehoux P;
Dusserget O, Chetouani F, Nedjar H, Glaser P, Kunst F, Cossart P;
Danielis J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;
Domínguez-Bernal G, Garrido-García P, Tíerrez-Martínez A, Amend A;
Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
Pérez-Díaz J, Baquero F, García Del Portillo F, Gómez-López N;
Pi Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
Pi Rose M, Voss H;
WIPI; 2002-010914/01.
Genomic sequence for Listeria monocytogenes, useful e.g., for treatment
and prevention of Listeria and related bacterial infections, and
related polypeptides
Claim 6; SEQ ID No 366; 192pp; French.
The present invention relates to the genome sequence of Listeria
monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
it are useful for selecting probes and primers for detecting fragments of
Listeria monocytogenes and related organisms, and for studying genetic
polymorphisms and other genomes. The present sequence is a protein
expressed by the genome sequence of the present invention. Proteins
expressed from the genome sequence are useful for raising specific
antibodies, identification of L. monocytogenes and related organisms, and
for biosynthesis and biodegradation, especially biosynthesis of Vitamin
B12. The genome sequence and proteins encoded by it are also useful for
selecting compounds that regulate gene expression and cell replication
and modulate L. monocytogenes-related diseases. In addition, the genome
sequence and proteins encoded by it are useful in pharmaceutical and
vaccine compositions for the treatment or prevention of infections by L.
monocytogenes and related organisms.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
Sequence 259 AA;
Query Match 63.8%; Score 852; DB 23; Length 259;
Best Local Similarity 63.4%; Pred. No. 2.4e-87;
Matches 161; Conservative 37; Mismatches 56; Indels 0; Gaps 0
Qy 1 LLLVIDGNTVTLGVYHDGKLEYHWRIETSRHKTEDEFGMLTSLFPHSGLMFQIDGI 60
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MLVIDGNTCTGVGYEKQLLKHWRMTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 60
Qy 61 IISVVPPIMFALERMCTKYFHIPEQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAATHL 120
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 IISVVPPIMHMETMTCVRYNIRPLIVPGIKTGLNLKVDNPRIGSDRIVNAVAASEE 120
Qy 121 YGNPLIVVDFGTATTCYVIDENKQVMGGAAPGITISTEALYSRAAKLPRIETRPDNI 180
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 YGTPVIVDFGTATTCYIDESGVYQGGALAPGIMISTEALYNRAAKLPVVDIASSQII 180
Qy 181 GKNTVSAMQSGILFYGVQGVGIVKRMQKQAKODLKVIATGGLAPLANESCDIVDPFF 240
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 GKSTVSSMQAGIFYGVQCEGIIAEMKKQSNASPVVATGTGLARMITEKSAVDILDPF 240
Qy 241 LTLKGLLELIYERNR 254
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 LTLKGLLELIYRRNK 254
RESULT 8
ID AAU91175 standard; Protein; 256 AA.
XX AC AAU91175;

```

XX	05-JUN-2002	(first entry)
XX	Pantothenate kinase (Coax) #13.	
XX	Pantothenate kinase; Coax; antibiotic; antimicrobial;	
XX	pantothenate kinase modulator; coenzyme A; bactericidal compound.	
XX	Clostridium difficile.	
XX	WC200216601-A2.	
XX	28-FEB-2002.	
XX	24-AUG-2001; 2001WO-US26531.	
XX	24-AUG-2000; 2000US-227860P.	
XX	20-MAR-2001; 2001US-0813453.	
XX	(OMNI-) OMNIGENE BIOPRODUCTS INC.	
XX	Yocum RR, Patterson TA;	
XX	WPI; 2002-269358/31.	
XX	N-PSDB; ABK54196.	
XX	Identifying potential antibiotic or antimicrobial agent, comprises	
XX	contacting composition comprising pantothenate kinase (Coax) protein	
XX	with test compound and identifying inhibitor of the Coax protein -	
XX	Claim 6; Page 105; 128pp; English.	
XX	The invention describes assays for identifying a (potential) antibiotic	
XX	comprising contacting an assay composition comprising a pantothenate	
XX	kinase (Coax) protein with a test compound, and determining the ability	
XX	of the test compound to inhibit the activity of the Coax protein, an	
XX	essential enzyme for the production of coenzyme A. Coax protein is a	
XX	valuable target for identifying bactericidal compounds. Coax modulating	
XX	agents can be used in an infectious animal model to determine the	
XX	efficacy, toxicity, or side effects of treatment with such an agent. This	
XX	is the amino acid sequence of a pantothenate kinase (Coax) protein	
XX	described in the invention.	
XX	Sequence 256 AA;	
XX	Query Match 59.5%; Score 794.5; DB 23; Length 256;	
XX	Best Local Similarity 60.8%; Pred. No. 7.4e-81;	
XX	Matches 155; Conservative 40; Mismatches 59; Indels 1; Gaps 1;	
Qy	1 LLLVIDVGNNTNTVLGVYHDKGLBYHWRISRKHTEDEFGMILRSLFDHSGLMPEQIDGI 60	
Db	1 MLLVDFVGNNTNMLVGIYKGDVLVYWRITDKRETSDEYGLISNLFYDYNVNISDIDVV 60	
Qy	61 IISVVPPINFALERMCTKYFHFIEPQIVGPMKTMGLNIKYDNPKEVGADRIVNAVAIHL 120	
Db	61 IISVVPPNVNMSLENFCIKYCKKQPLIVGPGIKTGLNIKYDNPKNQVGADRIVNAVAGIEK 120	
Qy	121 YGNPLIWDPGTATTTCYIDENQYWGGAAPGTTISTEALYSRAAKLPRIETRPNI 180	
Db	121 YGAPSLIVDFGTATTTCATSEKGEYLGGTAPGIKISSEALFOSASKLPRLVLAQPGMTI 180	
Qy	181 GKNTVSAMQSGILFGVYGVQVEGIVKRMKQAK-QDLKVIATGGTGLAPLIANESDCIDIVDP 239	
Db	181 CKSTVSAMQSGIYGVYGLVDKLIISMKKELNCDDVKVIATGGTGLAKLIASETSKIDYVDG 240	
Qy	240 FLTAKGLELIYERNR 254	
Db	241 FLTLEGLRIIYERKQ 255	
XX	RESULT 9	
XX	AAU91154	
XX	ID AAU91154 standard; Protein; 255 AA.	

```
XX AAU91154;
AC
XX
XX 05-JUN-2002 (first entry)
XX
XX Geobacter sulfurreducens pantothenate kinase Coax.
XX
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
XX Geobacter sulfurreducens.
OS
XX WO200216601-A2.
XX
XX 28-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-US26531.
XX
XX 24-AUG-2000; 2000US-227860P.
XX
XX 20-MAR-2001; 2001US-0813453.
XX
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
XX Yocum RR, Patterson TA;
XX
XX WPI; 2002-269358/31.
XX
XX Identifying potential antibiotic or antimicrobial agent, comprises
XX contacting composition comprising pantothenate kinase (Coax) protein
XX with test compound and identifying inhibitor of the Coax protein -
XX
XX Claim 10; Page 72-73; 128pp; English.
XX
XX The invention describes assays for identifying a (potential) antibiotic
XX comprising contacting an assay composition comprising a pantothenate
XX kinase (Coax) protein with a test compound, and determining the ability
XX of the test compound to inhibit the activity of the Coax protein, an
XX essential enzyme for the production of coenzyme A. Coax protein is a
XX valuable target for identifying bactericidal compounds. Coax modulating
XX agents can be used in an infectious animal model to determine the
XX efficacy, toxicity, or side effects of treatment with such an agent. This
XX is the amino acid sequence of a pantothenate kinase (Coax) protein
XX described in the invention.
XX
XX Sequence 255 AA;
XX
Query Match 56.8%; Score 756; DB 23; Length 255;
Best Local Similarity 57.5%; Pred. No. 1.6e-76;
Matches 146; Conservative 46; Mismatches 62; Indels 0; Gaps 0;
QY 1 LLLVIDVGNNTVLGVYHDKLEYHWRISTSRHKTDEFGMILRSFLDHSGLMFEQIDGI 60
DB 1 MLLVIDVGNNTVLGVYHDKLEYHWRISTSRHKTDEFGMILRSFLDHSGLMFEQIDGI 60
QY 61 IISVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRVNAV 120
DB 61 IISVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRVNAV 120
QY 121 YGNPLIVDPGTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETRPDNI 180
DB 121 YRTSLIIVDPGTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETRPDNI 180
QY 181 GKNVTSAMQSGILFGVYGVGVEIVKMKWQAKDLKVIATGGLAPLIANESDCIDVDP 240
DB 181 ARNTVNSMQAGIYGVYGLVDEIVTRMKAEKADAPRIATGGGLASLAPESKTEAVERY 240
QY 241 LTLKGLLEIYERNR 254
DB 241 LTLKGLLEIYERNR 254
```

RESULT 10
AAU91151

```
ID AAU91151 standard; Protein; 265 AA.
XX
XX AC AAU91151;
XX
XX DT 05-JUN-2002 (first entry)
XX
XX DE Streptomyces coelicolor pantothenate kinase Coax.
XX
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
XX OS Streptomyces coelicolor.
XX
XX PN WO200216601-A2.
XX
XX PD 28-FEB-2002.
XX
XX PF 24-AUG-2001; 2001WO-US26531.
XX
XX PR 24-AUG-2000; 2000US-227860P.
XX
XX PR 20-MAR-2001; 2001US-0813453.
XX
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
XX PI Yocum RR, Patterson TA;
XX
XX DR WPI; 2002-269358/31.
XX
XX XX Identifying potential antibiotic or antimicrobial agent, comprises
XX contacting composition comprising pantothenate kinase (Coax) protein
XX with test compound and identifying inhibitor of the Coax protein -
XX
XX PS Claim 10; Page 69-70; 128pp; English.
XX
XX CC The invention describes assays for identifying a (potential) antibiotic
XX comprising contacting an assay composition comprising a pantothenate
XX kinase (Coax) protein with a test compound, and determining the ability
XX of the test compound to inhibit the activity of the Coax protein, an
XX essential enzyme for the production of coenzyme A. Coax protein is a
XX valuable target for identifying bactericidal compounds. Coax modulating
XX agents can be used in an infectious animal model to determine the
XX efficacy, toxicity, or side effects of treatment with such an agent. This
XX is the amino acid sequence of a pantothenate kinase (Coax) protein
XX described in the invention.
XX
XX SQ Sequence 265 AA;
XX
Query Match 49.1%; Score 655; DB 23; Length 265;
Best Local Similarity 51.0%; Pred. No. 4.3e-65;
Matches 133; Conservative 46; Mismatches 74; Indels 8; Gaps 3;
QY 1 LLLVIDVGNNTVLGVYHDKLEYHWRISTSRHKTDEFGMILRSFLDHSGLMFEQ---- 56
DB 1 MLLVIDVGNNTVLGVYHDKLEYHWRISTSRHKTDEFGMILRSFLDHSGLMFEQ 60
QY 57 IDGIISVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRVNAV 115
DB 61 IDGIALCATVPSVHLREVTTRYGDPVAVLVEPGVKTGVPILTDHPKEVGADRVNAV 120
QY 116 AAHLIYGNPLIVDPGTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETR 175
DB 121 AAVELYGGPAIVDPGTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETR 180
QY 176 PNITGKNTVSMQSGILFGVYGVGVEIVKMKWQAKDLKVIATGGLAPLIANESD 232
DB 181 PRSIVGKNTVSMQSGILFGVYGVGVEIVKMKWQAKDLKVIATGGLAPLIANESD 240
QY 233 CIDIVDPPLTLKGLLEIYERN 253
DB 241 VIDEHEPMTLMLGLRLVIERN 261
```

RESULT 11

AAU91150
ID AAU91150 standard; Protein; 250 AA.
XX AC AAU91150;
XX AC AAU91150;
XX DT 05-JUN-2002 (first entry)
XX DE Clostridium acetobutylicum pantothenate kinase Coax.
XX DE Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Clostridium acetobutylicum.
XX PN WO200216601-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX PI Yocum RR, Patterson TA;
XX DR WPI; 2002-269358/31.
XX PT Identifying potential antibiotic or antimicrobial agent, comprises
XX PT contacting composition comprising pantothenate kinase (Coax) protein
XX PT with test compound and identifying inhibitor of the Coax protein -
XX PS Claim 10; Page 68-69; 128pp; English.
XX CC The invention describes assays for identifying a (potential) antibiotic
XX CC comprising contacting an assay composition comprising a pantothenate
XX CC kinase (Coax) protein with a test compound, and determining the ability
XX CC of the test compound to inhibit the activity of the Coax protein, an
XX CC essential enzyme for the production of coenzyme A. Coax protein is a
XX CC valuable target for identifying bactericidal compounds. Coax modulating
XX CC agents can be used in an infectious animal model to determine the
XX CC efficacy, toxicity, or side effects of treatment with such an agent. This
XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein
XX CC described in the invention.
XX SQ Sequence 250 AA;
Query Match 48.4%; Score 646.5; DB 23; Length 250;
Best Local Similarity 53.9%; Pred. No. 3.6e-64;
Matches 125; Conservative 46; Mismatches 56; Indels 5; Gaps 1;
QY 1 LLLVIDGNTNTVLGVHDGKLEVHWRIETSRHKTDEFGMLRSLFPHSGLMFEQIDGI 60
DB 18 VILVDVGNITVLGVYNDITLTAEWRLSTDLVRSADSYGIVQVNLFOQDKLPTLVGV 77
QY 61 IISVVPPIPFALERMCTKYFHIPEQIVGPGMKTGLNPKYDNPKVEVGADRVNVAAILH 120
DB 78 IISVVPNIMSLHEMIRKYFKNPLVVGPGIKTGINKYDNPKVEVGADRVNVAAILH 137
QY 121 YGNPLIVDFGTATTCYIDENKQYMGGAIAFGITISTEALYSRAAKLPRIETRPDNI 180
DB 138 YKRSIIIDFGTATTCFAVRENGDYLGAICPGIKVSSEALFERAAKLPVELIKPAYAI 197
QY 181 KNTVSAQSGILFGYGVQGVGIVKRMKQAKQDKL-----VIATGGIAPLI 227
DB 198 CKNTISSQSGIVRYLRQVLYLFEKLENLPDGRTRTSLVLTATGGIAPLI 249
RESULT 12
AAU91173
ID AAU91173 standard; Protein; 260 AA.
XX

AAU91173;
XX 05-JUN-2002 (first entry)
XX DE Pantothenate kinase (Coax) #11.
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Caulobacter crescentus.
XX PN WO200216601-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX PI Yocum RR, Patterson TA;
XX DR WPI; 2002-269358/31.
XX DR N-PSDB; ABK54194.
XX PT Identifying potential antibiotic or antimicrobial agent, comprises
XX PT contacting composition comprising pantothenate kinase (Coax) protein
XX PT with test compound and identifying inhibitor of the Coax protein -
XX PS Claim 10; Page 102-103; 128pp; English.
XX CC The invention describes assays for identifying a (potential) antibiotic
XX CC comprising contacting an assay composition comprising a pantothenate
XX CC kinase (Coax) protein with a test compound, and determining the ability
XX CC of the test compound to inhibit the activity of the Coax protein, an
XX CC essential enzyme for the production of coenzyme A. Coax protein is a
XX CC valuable target for identifying bactericidal compounds. Coax modulating
XX CC agents can be used in an infectious animal model to determine the
XX CC efficacy, toxicity, or side effects of treatment with such an agent. This
XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein
XX CC described in the invention.
XX SQ Sequence 260 AA;
Query Match 43.2%; Score 577; DB 23; Length 260;
Best Local Similarity 44.4%; Pred. No. 2.6e-56;
Matches 114; Conservative 50; Mismatches 91; Indels 2; Gaps 1;
QY 1 LLLVIDGNTNTVLGVHDGKLEVHWRIETSRHKTDEFGMLRSLFPHSGLMFEQIDGI 60
DB 1 MLIAEQGNTNTMFAIHHDGASVWAQWRSATSTTADYEYVWLSQLSMQGLGPRADAV 60
QY 61 IISVVPPIPFALERMCTKYFHIPEQIVGPGMKTGLNPKYDNPKVEVGADRVNVAAILH 120
DB 61 IISVVPQSIPLNLSRRYFNVEPLVIGENAKLIGDIVRIEKPSEAGADRLVNAIGAMV 120
QY 121 YGNPLIVDFGTATTCYIDENKQYMGGAIAFGITISTEALYSRAAKLPRIETRP--DN 178
DB 121 YPGPLVVIDSTATTTFDIVAADGAFEGGIIAPGINLSQALHEAAKLPRIAIQRPAGNR 180
QY 179 IIGKNTVSAMQSGILFGYGVQGVGIVKRMKQAKQDKLVKVIATGGLAPLINESCIDIVD 238
DB 181 IVGTDVTSAMQSGVFWGYISLIEGLVARIKAERGPMTVIATGTVASLFEATGDSIDHFD 240
QY 239 PFLTGLKGLIYERNRV 255
DB 241 SDLTIRGLLEIYRNTI 257
RESULT 13
ABP65945

Db 60 IISSTAPRVNLRNLCNRYFDCRPVYVKGFCCLPVAPRVDPGTTVPDRLVNTVAGYD 119
Qy 120 LYGNPLIIVDFGTATTCYCIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEITRPDNI 179
Db 120 RHGGDLIVDRGTATTPDVAPDGNAYIGVIAFGVNLSEALHMAAALPHVDVTKQGV 179
Qy 180 IGKNTVSAMQSGILFYGVGQVEGIVKRMKQAKDLKVIATGGIAPLIANESDCIDIYDP 239
Db 180 IGTNTVACIQSGVYMGYIGLVEGIVRQIRMERDRPMKVIATGGLASLFDLGLFDLFDKVED 239
Qy 240 FLTLKGLLEIYERNR 254
Db 240 DLTMHGLRLIFDYNK 254

RESULT 15
AAU91176
ID AAU91176 standard; Protein; 219 AA.
XX AC AAU91176;
XX DT 05-JUN-2002 (first entry)
XX DE Pantothenate kinase (Coax) #14.
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Dehalococcoides ethenogenes.
XX FN WO200216601-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX PI Yocum RR, Patterson TA;
XX WPI; 2002-269358/31.
XX N-PSDB; ABK54197.

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -
Claim 10; Page 106-107; 128pp; English.
The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

SQ Sequence 219 AA;
Query Match 38.6%; Score 515; DB 23; Length 219;
Best Local Similarity 45.8%; Pred. No. 2e-49;
Matches 97; Conservative 45; Mismatches 70; Indels 0; Gaps 0;
Qy 2 LLVIDVGNNTVILGVYHDKLEYHWRITESHKTEDEFGMILRSILFDHSGLMFQIDGII 61
Db 5 LVAVDIGNTSVNIIGIFEGEKLNNHGLSVAQRMADEYASLLGLLQHAGIHPELNKVI 64

Qy 62 ISSVVPIMPFALERMCTKYFHIEPQIVPGMKTGILNIKYDNPKEVGADRIVNAVAIHLV 121
Db 65 MCSVVPPLTTTTFEBVFKSYFKAAPLVVVGAGIKSGVKVRMDNPREVGADRIVNAVAARVLY 124
Qy 122 GNPLIIVDFGTATTCYCIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEITRPDNIIG 181
Db 125 PGACIIIVDMGTATTTDFDLSEGGAYIGGAIAPGIATSAQAIAEKTSKLPKIEIIRPAKVIG 184
Qy 182 KNTVSAMQSGILFYGVGQVEGIVKRMKQAKQ 213
Db 185 SNTVSAMQSGIYFGYIGLVEELVRRITQELGQ 216

Search completed: December 18, 2003, 13:01:37
Job time : 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2003, 12:54:22 ; Search time 31 Seconds
(without alignments)
1554.314 Million cell updates/sec

Title: US-09-813-453B-2

Perfect score: 1335

Sequence: 1 LLLVIDVGNNTVLGVYHDC.....PFLTLKGLLIYERNRGSV 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*

2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*

6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*

7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*

9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*

10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*

11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*

12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*

13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*

14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*

15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*

16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*

17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*

18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1335	100.0	258	10	US-09-813-453A-2
2	1114	83.4	233	10	US-09-813-453A-17
3	1046	78.4	258	10	US-09-813-453A-49
4	1034	77.5	262	10	US-09-813-453A-45
5	987	73.9	254	10	US-09-813-453A-47
6	794.5	59.5	256	10	US-09-813-453A-55
7	756	56.6	255	10	US-09-813-453A-7
8	655	49.1	265	10	US-09-813-453A-4
9	655	49.1	265	15	US-10-156-761-12224
10	646.5	48.4	250	10	US-09-813-453A-3
11	577	43.2	260	10	US-09-813-453A-51
12	537	40.2	258	10	US-09-813-453A-6
13	515	38.6	219	10	US-09-813-453A-57
14	494.5	37.0	272	10	US-09-813-453A-5
15	493.5	37.0	272	10	US-09-712-363-276

16	428	32.1	262	10	US-09-813-453A-8	Sequence 8, Appli
17	405.5	30.4	246	10	US-09-813-453A-9	Sequence 9, Appli
18	338.5	25.4	212	10	US-09-813-453A-59	Sequence 59, Appli
19	327.5	24.5	273	10	US-09-813-453A-10	Sequence 10, Appli
20	319	23.9	257	10	US-09-813-453A-53	Sequence 53, Appli
21	271.5	20.3	262	10	US-09-813-453A-11	Sequence 11, Appli
22	207	15.5	244	10	US-09-813-453A-41	Sequence 41, Appli
23	203	15.2	241	10	US-09-813-453A-63	Sequence 63, Appli
24	163	12.2	249	10	US-09-813-453A-70	Sequence 70, Appli
25	163	12.2	257	10	US-09-813-453A-13	Sequence 13, Appli
26	154.5	11.6	229	10	US-09-813-453A-12	Sequence 12, Appli
27	154	11.5	249	10	US-09-813-453A-61	Sequence 61, Appli
28	151	11.3	460	10	US-09-813-453A-39	Sequence 39, Appli
29	150	11.2	592	10	US-09-813-453A-22	Sequence 22, Appli
30	150	11.2	592	10	US-09-813-453A-43	Sequence 43, Appli
31	150	11.2	592	12	US-10-320-800-48	Sequence 48, Appli
32	134.5	10.1	242	10	US-09-813-453A-65	Sequence 65, Appli
33	133	10.0	248	10	US-09-813-453A-20	Sequence 20, Appli
34	129.5	9.7	267	10	US-09-813-453A-15	Sequence 15, Appli
35	109.5	8.2	223	10	US-09-895-913A-74	Sequence 74, Appli
36	109.5	8.2	223	10	US-09-813-453A-14	Sequence 14, Appli
37	109.5	8.2	223	10	US-09-813-453A-67	Sequence 67, Appli
38	109	8.2	209	10	US-09-813-453A-21	Sequence 21, Appli
39	80	6.0	449	9	US-09-815-242-5474	Sequence 5474, Ap
40	80	6.0	449	9	US-09-815-242-12348	Sequence 12348, A
41	80	6.0	449	9	US-09-815-242-12793	Sequence 12793, A
42	79	5.9	636	9	US-09-205-658-160	Sequence 160, App
43	79	5.9	636	12	US-09-963-693-160	Sequence 160, App
44	79	5.9	923	15	US-10-262-538-6	Sequence 6, Appli
45	78.5	5.9	337	15	US-10-075-846-12	Sequence 12, Appli

ALIGNMENTS

RESULT 1

US-09-813-453A-2

; Sequence 2, Application US/09813453A

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; TITLE OF INVENTION: ANTIBIOTICS

; FILE REFERENCE: CGZ-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: 09/667,569

; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 258

; TYPE: PRT

; ORGANISM: Bacillus subtilis

US-09-813-453A-2

Query Match	100.0%;	Score 1335;	DB 10;	Length 258;			
Best Local Similarity	100.0%;	Pred. No. 1.6e-144;					
Matches 258;	Conservative	0;	Mismatches	0; Indels 0; Gaps 0;			
Qy	1	LLLVLDVGNNTVLGVYHDC	KLKYHWR	IFTSRKHTDEFGMILRS	LDHSLGLMPEQIDGI	60	
Db	1	LLLVLDVGNNTVLGVYHDC	KLKYHWR	IFTSRKHTDEFGMILRS	LDHSLGLMPEQIDGI	60	
Qy	61	IISVVVPPINFALERMCTKY	FHLEPQIVG	PMKTKGLN	KYDNPKEVCADR	IVNAVAIHL	120
Db	61	IISVVVPPINFALERMCTKY	FHLEPQIVG	PMKTKGLN	KYDNPKEVCADR	IVNAVAIHL	120
Qy	121	YGNPLIVDFGTATTTCY	IDENKQYMG	GAIPGTTISTEALYSRAAKLP	RIETITRPDNI	180	

Db 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIEITRPDNI 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-09-813-453A-49

Qy 181 GKNTVSAMQSGILFGYVGVGVEGIVKRMKQAKODLKVIATGGLAPLIANESDCIDVDPF 240
Db 181 GKNTVSAMQSGILFGYVGVGVEGIVKRMKQAKODLKVIATGGLAPLIANESDCIDVDPF 240
Qy 241 LTLKGLLELIYERNVGSV 258
Db 241 LTLKGLLELIYERNVGSV 258

RESULT 2
US-09-813-453A-17
; Sequence 17, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-813-453A-17

Query Match 83.4%; Score 1114; DB 10; Length 233;
Best Local Similarity 99.5%; Pred. No. 3.2e-119;
Matches 213; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLLVIDVGNNTVLGVYHDKLEYHWRITSRHKTDEFGMLIRSLFDHSGLMFEQIDGI 60
Db 1 LLLVIDVGNNTVLGVYHDKLEYHWRITSRHKTDEFGMLIRSLFDHSGLMFEQIDGI 60
Qy 61 ITSSVVPPIFALERMCTKYFHIPOIVGPGMKTGLNIDNPKEVGADRIVNAVAIHL 120
Db 61 ITSSVVPPIFALERMCTKYFHIPOIVGPGMKTGLNIDNPKEVGADRIVNAVAIHL 120
Qy 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIEITRPDNI 180
Db 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIEITRPDNI 180
Qy 181 GKNTVSAMQSGILFGYVGVGVEGIVKRMKQAKOD 214
Db 181 GKNTVSAMQSGILFGYVGVGVEGIVKRMKQAKOD 214

RESULT 3
US-09-813-453A-49
; Sequence 49, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-09-813-453A-49

Query Match 78.4%; Score 1046; DB 10; Length 258;
Best Local Similarity 78.3%; Pred. No. 2.3e-111;
Matches 198; Conservative 32; Mismatches 23; Indels 0; Gaps 0;

Qy 1 LLLVIDVGNNTVLGVYHDKLEYHWRITSRHKTDEFGMLIRSLFDHSGLMFEQIDGI 60
Db 1 MIFVLVDGNTNTVLGVYDGDDELKHHWRITSRKTSKTEDEYGMKIKALLNHVGLQPSDIRGI 60
Qy 61 ITSSVVPPIFALERMCTKYFHIPOIVGPGMKTGLNIDNPKEVGADRIVNAVAIHL 120
Db 61 ITSSVVPPIFALERMCTKYFHIPOIVGPGMKTGLNIDNPKEVGADRIVNAVAIHL 120
Qy 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIEITRPDNI 180
Db 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIEITRPDNI 180
Qy 181 GKNTVSAMQSGILFGYVGVGVEGIVKRMKQAKODLKVIATGGLAPLIANESDCIDVDPF 240
Db 181 GKNTVSAMQSGILFGYVGVGVEGIVSRMKAKSKI PPKVIATGGLAPLIASESDIIDVDPF 240
Qy 241 LTLKGLLELIYERN 253
Db 241 LTLKGLLELYEKN 253

RESULT 4
US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-09-813-453A-45

Query Match 77.5%; Score 1034; DB 10; Length 262;
Best Local Similarity 75.1%; Pred. No. 5.7e-110;
Matches 190; Conservative 34; Mismatches 29; Indels 0; Gaps 0;

Qy 1 LLLVIDVGNNTVLGVYHDKLEYHWRITSRHKTDEFGMLIRSLFDHSGLMFEQIDGI 60
Db 1 MIFVLVDGNTNTVLGVYFEGELRQHWRMETDRHKTDEYGLVKOLLEHGLSFDVKGI 60
Qy 61 ITSSVVPPIFALERMCTKYFHIPOIVGPGMKTGLNIDNPKEVGADRIVNAVAIHL 120
Db 61 ITSSVVPPIFALERMCTKYFHIPOIVGPGMKTGLNIDNPKEVGADRIVNAVAIHL 120
Qy 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIEITRPDNI 180
Db 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIEITRPDNI 180
Qy 181 GKNTVSAMQSGILFGYVGVGVEGIVKRMKQAKODLKVIATGGLAPLIANESDCIDVDPF 240

Db 181 GNTVSAMQSGILYGVQVGVGIVKRMKEAKQEPKVIATGGLAKLISEESNVIVIDVDPF 240
Qy 241 LTLKGLLEIYERN 253
Db 241 LTLKGLYLYERN 253

RESULT 5
US-09-813-453A-47
; Sequence 47, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-09-813-453A-47

Query Match 73.9%; Score 987; DB 10; Length 254;
Best Local Similarity 72.3%; Pred. No. 1.3e-104;
Matches 183; Conservative 33; Mismatches 37; Indels 0; Gaps 0;
Qy 1 LLLVIDGNTNTVLGVHDGKLEYHWRIETSRHKTEDEFGMLRSLFDHSGLMFEQIDGI 60
Db 1 MLVIDGNTNTVLGVQDETIVHHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQDIDGI 60
Qy 61 IISVVPPIPALERMCTKYFHIPOIVGPMKGTGLNIKYDNPKVEVGADRIVNAVAHL 120
Db 61 VISSVPPMPSLEQMCKKYFHVTPMIIIGPKITGLNIKYDNPKVEVGADRIVNAVAEL 120
Qy 121 YGNPLIVVDGTTATTCYVIDENKQYMGGAIAPIGTISTEALYSRAAKLPRIETRPDNI 180
Db 121 YGPAIVVDGTTATTCYCLINEKKQYAGGVIAPIGTISTEALYHRASKLPRIETRAKQV 180
Qy 181 GNTVSAMQSGILFYGVQVGVGIVKRMKQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
Db 181 GTNTIDSMQSGIFYGVQVGVGIVKRMKQAKQAESEPKVIATGGLAKLIGTSETIDVIDSF 240
Qy 241 LTLKGLLEIYERN 253
Db 241 LTLKGLQIYKKN 253

RESULT 6
US-09-813-453A-55
; Sequence 55, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-09-813-453A-55
Query Match 59.5%; Score 794.5; DB 10; Length 256;
Best Local Similarity 60.8%; Pred. No. 1.7e-82;
Matches 155; Conservative 40; Mismatches 59; Indels 1; Gaps 1;
Qy 1 LLLVIDGNTNTVLGVHDGKLEYHWRIETSRHKTEDEFGMLRSLFDHSGLMFEQIDGI 60
Db 1 MLLVFDGNTNMVLGIYKDKLVNWKRIKTDREKTSDEYGLISNLFYDNNVISDIDDV 60
Qy 61 IISVVPPIPALERMCTKYFHIPOIVGPMKGTGLNIKYDNPKVEVGADRIVNAVAHL 120
Db 61 IISVWPNVMHSLNFICIKYCKQPLIVGPKITGLNIKYDNPKQVGADRIVNAVAGIEK 120
Qy 121 YGNPLIVVDGTTATTCYVIDENKQYMGGAIAPIGTISTEALYSRAAKLPRIETRPDNI 180
Db 121 YGAPSLVVDGTTATTCFCAISEKGEYLGTTAPGKIKISEALFQSAKLPVELAKPGMTI 180
Qy 181 GNTVSAMQSGILFYGVQVGVGIVKRMKQAK-QDLKVIATGGLAPLIANESDCIDIVDP 239
Db 181 CKSTVSAMQSGIIVGYVGLVDKIISIMKELNCDVVKVIATGGLAKLIASETKSIDYVDG 240
Qy 240 FTLKGLLEIYERNR 254
Db 241 FTLLEGLRIIYKKNQ 255

RESULT 7
US-09-813-453A-7
; Sequence 7, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Geobacter sulfurreducens
US-09-813-453A-7

Query Match 56.6%; Score 756; DB 10; Length 255;
Best Local Similarity 57.5%; Pred. No. 4.3e-78;
Matches 146; Conservative 46; Mismatches 62; Indels 0; Gaps 0;
Qy 1 LLLVIDGNTNTVLGVHDGKLEYHWRIETSRHKTEDEFGMLRSLFDHSGLMFEQIDGI 60
Db 1 MLLVIDGNTNTVLGIYDGERLVDNRVSTDKARTTDEYGLINELFRLAGLQDITRAV 60
Qy 61 IISVVPPIPALERMCTKYFHIPOIVGPMKGTGLNIKYDNPKVEVGADRIVNAVAHL 120
Db 61 IISVWPPLTGVLRLSLGYFGMRPLVVGPKITGMPIQYDNPREVGADRIVNAVAGYK 120
Qy 121 YGNPLIVVDGTTATTCYVIDENKQYMGGAIAPIGTISTEALYSRAAKLPRIETRPDNI 180
Db 121 YRTSLIIVDGTATTTFDYVNRKGEYCGGAIAPIGLVISTEALFORASKLPYVDIIRPSAI 180

```

QY 181 GKNTVSAMQSGILFGYVGVGKLEKMKWQAKQDLKVIATGATGLAPLANESDCIDIVDPF 240
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 181 ARNTVSMQAGIYGYVGLVDEIVTRKAEKADPRVIATGGLASIAESKTIIEVEY 240
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 241 LTLKGLLEIYERNR 254
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 241 LTLLEGLIYERNR 254
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

```

RESULT 8
US-09-813-453A-4
; Sequence 4, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Streptomyces coelicolor
US-09-813-453A-4

```

```

Query Match 49.1%; Score 655; DB 10; Length 265;
Best Local Similarity 51.0%; Pred. No. 1.8e-66;
Matches 133; Conservative 46; Mismatches 74; Indels 8; Gaps 3;

```

```

QY 1 LLLVDVGNNTVLGVYHDGKLEKLEKMKWQAKQDLKVIATGATGLAPLANESDCIDIVDPF 56
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1 MLTIDVGNTHTVLGLFDGEDIVHEHWRISTDSRTADELAVLQGLMGHPLLGDELGDG 60
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 57 IDGIISSVVPPIMFALERMCTKYFHIPOI-VGPGMKTGLNPKYDNPKEVGADRIINAV 115
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 61 IDGIAICATVPSVLHELRETVRYGDPVAVLVEPGIKTGVPLMDNPKYDNPKEVGADRIINAV 120
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 116 AAHLVGNPLIVVDFGTATTCYVIDENKQYMGGAIAPIGTITSTEALYSRAAKLPRIEITR 175
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 121 AAVELYGGPAIVVDFGTATTFDAVSARGEYIGVIAPIGIEISVEALGVKAQLRKIEVAR 180
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 176 PDNIIGKNTVSAMQSGILFGYVGVGKLEKMKWQAKQDLKVIATGATGLAPLANESD 232
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 181 PRSVIGKNTVEAMQSGIYVGFAGQVGVVNRMAELADDDPDDVTVIATGGLAPMVLGESS 240
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 233 CIDIVDPFLTKGLELIYERN 253
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 241 VIDEHEPWLTLGLRLVYERN 261
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

```

RESULT 9
US-10-156-761-12224
; Sequence 12224, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761

```

```

; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12224
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12224

```

```

Query Match 49.1%; Score 655; DB 15; Length 265;
Best Local Similarity 51.7%; Pred. No. 1.8e-66;
Matches 135; Conservative 42; Mismatches 76; Indels 8; Gaps 3;

```

```

QY 1 LLLVDVGNNTVLGVYHDGKLEKLEKMKWQAKQDLKVIATGATGLAPLANESDCIDIVDPF 56
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1 MLTIDVGNTHTVLGLFDGEDIVHEHWRISTDSRTADELAVLQGLMGHPLLGDELGDG 60
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 57 IDGIISSVVPPIMFALERMCTKYFHIPOI-VGPGMKTGLNPKYDNPKEVGADRIINAV 115
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 61 IDGIAICATVPSVLHELRETVRYGDPVAVLVEPGIKTGVPLMDNPKYDNPKEVGADRIINAV 120
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 116 AAHLVGNPLIVVDFGTATTCYVIDENKQYMGGAIAPIGTITSTEALYSRAAKLPRIEITR 175
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 121 AAVELYGGPAIVVDFGTATTFDAVSARGEYIGVIAPIGIEISVEALGVKAQLRKIEVAR 180
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 176 PDNIIGKNTVSAMQSGILFGYVGVGKLEKMKWQAKQDLKVIATGATGLAPLANESD 232
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 181 PRSVIGKNTVEAMQSGIYVGFAGQVGVVNRMAELADDDPDDVTVIATGGLAPMVLGESS 240
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 233 CIDIVDPFLTKGLELIYERN 253
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 241 VIDEHEPWLTLGLRLVYERN 261
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

```

RESULT 10
US-09-813-453A-3
; Sequence 3, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-09-813-453A-3

```

```

Query Match 48.4%; Score 646.5; DB 10; Length 250;
Best Local Similarity 53.9%; Pred. No. 1.5e-65;
Matches 125; Conservative 46; Mismatches 56; Indels 5; Gaps 1;

```

```

QY 1 LLLVDVGNNTVLGVYHDGKLEKLEKMKWQAKQDLKVIATGATGLAPLANESDCIDIVDPF 60
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 18 VILVDVGNNTVLGVYHDGKLEKLEKMKWQAKQDLKVIATGATGLAPLANESDCIDIVDPF 77
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 61 IISVVPPIMFALERMCTKYFHIPOI-VGPGMKTGLNPKYDNPKEVGADRIINAVAAIHL 120
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 78 IISVVPPIMFALERMCTKYFHIPOI-VGPGMKTGLNPKYDNPKEVGADRIINAVAAIHL 137
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```


	Query Match	38.6%	Score 515;	DB 10;	Length 219;
	Best Local Similarity	45.8%;	Pred. No. 1.5e-50;		
	Matches	97;	Conservative 45;	Mismatches 70;	Indels 0; Gaps 0;
Qy	2	LLVIDGNTNTVLGVYHDKLELVHRIETSRHKTEDEFGMILERSLPDHSGLMFEQIDGII	61		
Db	5	LVAVDIGNTSVNIIGIFEGBKLLANMHLGVSQAORMDEYASLLGLQLQAHGIIPEELNRVI	64		
Qy	62	ISSWPPPIPMFALERMCTKYPHTEPQIVGPGMTGLNIKVDNPKVEGADRIVNAAVAIHLV	121		
Db	65	MCSVPEPLTTTPEEVFKSVFKAAPLVVVGAGISGVKVRMDNPVEVGADRIVNAAAARVLY	124		
Qy	122	GNPLIVDFGTATTCYVIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETRPNNIIG	181		
Db	125	PGACIIVDMGTATTFOTLSEGGAYIGCAIAPGATSAOAIAEKTSKLPKEIIRPAKVIG	184		

```

; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 276
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-712-363-276

Query Match          37.0%; Score 493.5; DB 10; Length 272;
Best Local Similarity 38.8%; Pred. NO. 6.2e-48;
Matches 102; Conservative 55; Mismatches 93; Indels 13; Gaps 4;

QY      1  LLLIVDVGNVTNVLGVY----HDGKLEYHWRTSRRHKTEDEFGMLRSLRFDHSGLMFEQ 56
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       1  VLLAIDVRNTHTVVGLSGMKEHAKVVOQWRIRTESEVTADELALTYDGLIGEDS---ER 57

QY      57  IDGIIISVVPPIMPFALEMRMCTKYPIHEPOI-VGPGMKTGLNICYDNPKVEVGADRIVNAV 115
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       58  LGTGAALSTVSPVLHVEVRIMLQOYWPSPVHLIEPVRGTGIPLLVDNPKVEVGADRIVNCL 117

QY      116  AAHLHYGNLIVDFGTATTATCYCIDENKQYMGGAIAPIGTITISTEALYSRAAKLPRIEITR 175
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       118  AAYDRFKAAIVDFGSSICVDVWSAKGEFLGGAIAPGVQVSSDAAARSAAALRRVELAR 177

QY      176  PDNIIGKNTVSAMQSILFYGVQVEGIVKRMK-----WQAKODLKVIATGGLAPLIANE 230
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       178  PRSVWGNKTVECNQAGAVFGFAGLVLDGLVGRITREDSVGSFVDHDAIVATGHTAPLLPE 237

QY      231  SDCIDIVDPFLTLKGLLEIVERN 253
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db       238  LHTVDHYDQHLTLQGLRLYVERN 260
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

Search completed: December 18, 2003, 13:00:43
Job time : 32 secs

```

QY 182 KNTVSAQSGILFGYVGQVEGIVKRMKWQAQK 213
| | | | | | | | | | | : | : | : |
DB 185 SNTVSAQSGIYFGYIGLVEELVRRIQTGLQG 216

RESULT 14
US-09-813-453A-5
; Sequence 5, Application US/09813453A
; Patent No. US20020168681A1.
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF ANTIBIOTICS
; FILE REFERENCE: CGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 5
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-453A-5

Query Match 37.0%; Score 494.5; DB 10; Length 272;
Best Local Similarity 38.8%; Pred. No. 4.8e-48;
Matches 102; Conservative 55; Mismatches 93; Indels 13; Gaps 4;

QY 1 LLLVIDGVNNTVLGVY----HCKLEYHWRITSRHKTEDEFGMILRSFLDHSGLMFEO 56
DB 1 MLLAIIDVRNTHTVVGLLSGMKEHAKVVQQWRIETSEVTADALATTDGLIGDS---ER 57
QY 57 IDGIISVVPPIMFALERCTKYFHIEPOI-VCPGMKTGTINIKYDNPKVGGADRVNAV 115
DB 58 LTGTAAUUVPSVLHEVRIMLDQWPSPVHLIEFGVTRTGLPLVDNPKEVGADRIVNCL 117
QY 116 AAHLNGNLIVDFGTATTTCYIDENKOYMGGIAFPGITISTEALYSRAAKLPRIETR 175
DB 118 AAYDRPFAAIVDFCGSSICVDVVSAGKEFLGGAIFGVQSSDAARAARLRVELAR 177
QY 176 PDNIIGNTVSAMQSGLIFGVQVEGIVKRMK-----WQAKQDLKVATGGLAPLANE 230
DB 178 PRSVGVKNVTCECMAQAGAVFGAGLVLDGLVGRIREDDVSGFSVDHDVAIVAATGHTAPLLPE 237
QY 231 SDCIDVDVPFLTKLGLELIVERN 253
DB 238 LHTVDHYDQHLLTQLGLRLVERN 260

RESULT 15
US-09-712-363-276
; Sequence 276, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2003, 13:00:04 ; Search time 21 Seconds
(without alignments)
1181.501 Million cell updates/sec

Title: US-09-813-453B-2

Perfect score: 1335

Sequence: 1 LLLVIDVGNNTVTLGVYHGD.....PFLTLKGLLEIYERNRGSV 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1114	83.4	233	2	S66100 conserved hypotet
2	987	73.9	254	2	S63660 hypotet
3	852	63.8	259	2	AF1102 conserved hypotet
4	847	63.4	259	2	AF1484 conserved hypotet
5	775	58.1	273	2	E97293 probable transcrip
6	655	49.1	265	2	T36391 hypotet
7	577	43.2	261	2	B87489 transcription acti
8	519.5	38.9	274	2	H86937 conserved hypotet
9	494.5	37.0	272	2	A70955 hypotet
10	428	32.1	262	2	E75516 conserved hypotet
11	405.5	30.4	246	2	D72320 conserved hypotet
12	327.5	24.5	273	2	D71326 conserved hypotet
13	271.5	20.3	262	2	F70165 conserved hypotet
14	170	12.7	276	2	A12292 hypotet
15	163	12.2	257	2	E75539 hypotet
16	154.5	11.6	229	2	E70455 hypotet
17	150	11.2	292	2	H81009 BirA protein/Bvg a
18	150	11.2	592	2	A82637 probable biotin-la
19	134.5	10.1	242	2	A82637 conserved hypotet
20	133	10.0	248	2	H83111 hypotet
21	129.5	9.7	267	2	I40327 bai protein - Bord
22	117	8.8	224	2	A99571 conserved hypotet
23	116.5	8.7	223	2	G71887 hypotet
24	109.5	8.2	223	2	F64627 hypotet
25	109	8.2	209	2	H81382 hypotet
26	104.5	7.8	597	2	B69231 probable electon
27	95	7.1	467	2	AB3142 hypotet
28	98145	7.1	520	2	G98145 probable aminotran
29	93	7.0	1014	2	H71602 protein with DnaJ

30 91.5 6.9 828 2 G82583
31 90.5 6.8 452 1 YBBSG2
32 88 6.6 256 1 H83002
33 87.5 6.6 287 2 D90540
34 87 6.5 336 2 AE0625
35 86.5 6.5 205 2 AC0612
36 86.5 6.5 205 2 AG0680
37 86.5 6.5 243 2 D64300
38 86 6.4 336 1 S13824
39 85 6.4 336 1 DEEDCO
40 85 6.4 336 2 E90757
41 85 6.4 336 2 E96792
42 84.5 6.3 317 2 E96792
43 84 6.3 443 1 JCS298
44 83.5 6.3 205 2 B85615
45 83.5 6.3 205 2 D90751

ALIGNMENTS

RESULT 1

S66100

Conserved hypothetical protein yacB - Bacillus subtilis

C;Species: Bacillus subtilis

C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000

C;Accession: S66100; E69740

R;Ogasawara, N.; Nakai, S.; Yoshikawa, H.

DNA Res. 1, 1-14, 1994

A;Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chrom

A;Reference number: S65967; MUID:96051385; PMID:7584024

A;Accession: S66100

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-233 <OGA>

A;Cross-references: EMBL:D26185; NID:G467326; PIDN:BAA05305.1; PID:di005847; PID:g467459

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993

R;Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero

C.; Bron, S.; Brullet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallert

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,

Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Sadaie, Y.; Sato, T.; Scanlon,

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seror,

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A;Reference number: A69580; MUID:98044033; PMID:9384377

A;Accession: E69740

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-233 <KUN>

A;Cross-references: GB:Z99104; GB:AL009126; NID:G2632267; PIDN:CAB11846.1; PID:e1182003;

A;Experimental source: strain 168

C;Genetics:

A;Gene: yacB

A;Start codon: TTG

C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 83.4%; Score 1114; DB 2; Length 233;
Best Local Similarity 99.5%; Pred. No. 3.1e-88;
Matches 213; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LLLVIDVGNNTVTLGVYHGDLEHYHRIETSRHKTDEFCGMILRSFDHSGLMFEQIDGI 60

Db 1 LLLVIDVGNNTVTLGVYHGDLEHYHRIETSRHKTDEFCGMILRSFDHSGLMFEQIDGI 60

Oy 61 IISVVPPIFMALERMCTKYFHIEPQIVGPMKTGLNPKYDNPKEVGADRIVNAVAIHL 120

Db 61 IISVVPPIMFALERMCTKYFHIEPOIVGPGMKTGLNICYNDPKEVGADRIVNAVAAILH 120
 QY 121 YGNPLIVDPFGTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETRPDNI 180
 Db 121 YGNPLIVDPFGTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETRPDNI 180
 QY 181 GKNTVSAMQSGILFGVVGQVEGIVKRMKQAKOD 214
 Db 181 GKNTVSAMQSGILFGVVGQVEGIVKRMKQAKOD 214
 RESULT 2
 F83660
 hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: F83660
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: F83660
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-254 <STO>
 A:Cross-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BA03805.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH0086
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
 Query Match 73.9%; Score 987; DB 2; Length 254;
 Best Local Similarity 72.3%; Pred. No. 2.8e-77;
 Matches 183; Conservative 33; Mismatches 37; Indels 0; Gaps 0;
 QY 1 LLLVIDVGNNTVLGVYHDGKLEHYHRIETSRHKTDEFGMILRSFLDHSGLMFEQIDGI 60
 Db 1 MILVIDVGNNTVLGVYQDETLLVHHRLATSRKQTEDEVAMTVRSFLDHAGLQFQIDGI 60
 QY 61 IISVVPPIMFALERMCTKYFHIEPOIVGPGMKTGLNICYNDPKEVGADRIVNAVAAILH 120
 Db 61 VISSVPPMFMFSEQCKYFHYFTPEIIGPIKTKGLNICYNDPKEVGADRIVNAVAAILH 120
 QY 121 YGNPLIVDPFGTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETRPDNI 180
 Db 121 YGYPAIVDPFGTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETRPDNI 180
 QY 181 GKNTVSAMQSGILFGVVGQVEGIVKRMKQAKODLKVIATGGLAPLIANESDCIDIVDPF 240
 Db 181 GTNTIISMQSGIFGVVQVDGVVKKMKAQSEPKVIATGGLAKLIGTSETIDVDSF 240
 QY 241 LTLKGLLELIYERN 253
 Db 241 LTLKGLLIYKKN 253
 RESULT 3
 AF1102
 conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain EGD-e
 C:Species: Listeria monocytogenes
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
 C:Accession: AF1102
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
 D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
 Science 294, 849-852, 2001
 A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
 A:Title: Comparative Genomics of Listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AF1102
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-259 <GLA>
 A:Cross-references: GB:NC_003210; PIDN:CAD00748.1; PID:g16409586; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: lmo0221
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
 Query Match 63.8%; Score 852; DB 2; Length 259;
 Best Local Similarity 63.4%; Pred. No. 1.1e-65;
 Matches 161; Conservative 37; Mismatches 56; Indels 0; Gaps 0;
 QY 1 LLLVIDVGNNTVLGVYHDGKLEHYHRIETSRHKTDEFGMILRSFLDHSGLMFEQIDGI 60
 Db 1 MILVIDVGNNTVLGVYEQKLLKHWMTTDRHRTSDELGMTVLNFFSVANLTSPDIQGI 60
 QY 61 IISVVPPIMFALERMCTKYFHIEPOIVGPGMKTGLNICYNDPKEVGADRIVNAVAAILH 120
 Db 61 IISVVPPIMHAMETMTCVRYFNIRPLIVGPGIKTKGLNKLKVDNPREIGSDRIVNAVAASEE 120
 QY 121 YGNPLIVDPFGTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETRPDNI 180
 Db 121 YGTPVIVDPFGTATTCYIDEAGVYQGGAIAPGIMISTEALYNRAAKLPVDIAESSQII 180
 QY 181 GKNTVSAMQSGILFGVVGQVEGIVKRMKQAKODLKVIATGGLAPLIANESDCIDIVDPF 240
 Db 181 GKSTVSSMQAGIFGVGQCEGIIAEMKKQSNASPVVATGGLARMIKSSNAVDILDPF 240
 QY 241 LTLKGLLELIYERN 254
 Db 241 LTLKGLLELYRENK 254
 RESULT 4
 AF1464
 conserved hypothetical protein lin0253 [imported] - Listeria innocua (strain Clip11262)
 C:Species: Listeria innocua
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
 C:Accession: AF1464
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
 D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
 Science 294, 849-852, 2001
 A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
 A:Title: Comparative Genomics of Listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AF1464
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-259 <GLA>
 A:Cross-references: GB:AL592022; PIDN:CAC95486.1; PID:g16412682; GSPDB:GN00178
 A:Experimental source: strain Clip11262
 C:Genetics:
 A:Gene: lin0253
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
 Query Match 63.4%; Score 847; DB 2; Length 259;
 Best Local Similarity 62.6%; Pred. No. 3e-65;
 Matches 159; Conservative 39; Mismatches 56; Indels 0; Gaps 0;
 QY 1 LLLVIDVGNNTVLGVYHDGKLEHYHRIETSRHKTDEFGMILRSFLDHSGLMFEQIDGI 60
 Db 1 MILVIDVGNNTVLGVYEQKLLKHWMTTDRHRTSDELGMTVLNFFSVANLTSPDIQGI 60
 QY 61 IISVVPPIMFALERMCTKYFHIEPOIVGPGMKTGLNICYNDPKEVGADRIVNAVAAILH 120
 Db 61 IISVVPPIMHAMETMTCVRYFNIRPLIVGPGIKTKGLNKLKVDNPREIGSDRIVNAVAASEE 120
 QY 121 YGNPLIVDPFGTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETRPDNI 180
 Db 121 YGTPVIVDPFGTATTCYIDEAGVYQGGAIAPGIMISTEALYNRAAKLPVDIAESSQII 180

Qy 181 GKNTVSAMQSGILFGYGVQVEGIVKRMKWAQKDLKVIATGGLAPLIANESDCIDIVDPF 240
 Db 181 GKSTVASHQAQIFGYGFCQCEGIIAEMKKQNTSPVVVATGGLARMITEKSSAVDILDPF 240

Qy 241 LTLKGLLELYERNR 254
 Db 241 LTLKGLLELYERNR 254

RESULT 5
 E97293
 Probable transcription regulator, homolog of Bvg accessory factor [imported] - Clostridi
 C:Species: Clostridium acetobutylicum
 C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
 C:Accession: E97293
 R:Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
 J. Bacteriol. 183, 4823-4836, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: E97293
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-273 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:gl5026270; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC3200
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 58.1%; Score 775; DB 2; Length 273;
 Best Local Similarity 56.4%; Pred. No. 4.8e-59;
 Matches 146; Conservative 55; Mismatches 54; Indels 4; Gaps 1;

Qy 1 LLLVIDGNTNTVLGVYHDKLEWHRIETSRHKTDEFGMLRSLFDHSGLMFEQIDGI 60
 Db 12 VILVDGNTNVLGVYHDKLEWHRIETSRHKTDEFGMLRSLFDHSGLMFEQIDGI 71

Qy 61 IISVVPPIMFALERMCTKYFHIEPOIVGPMKTGLNPKYDNPKEVGADRIVNAVAIHL 120
 Db 72 IISVVPPIMFALERMCTKYFHIEPOIVGPMKTGLNPKYDNPKEVGADRIVNAVAIHL 131

Qy 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETRP 180
 Db 132 YKRSLLIIDFTATTFCVAVRENGDYLGCAICPGIKVSSEALFEKAALPRVELIKPAYAI 191

Qy 181 GKNTVSAMQSGILFGYGVQVEGIVKRMKWAQKDLK----VIATGGLAPLIANESDCIDI 236
 Db 192 CKNTIISQSGIVGYIGQVRYIVRMKEBELQEEKEPELVATGGLAKLISEAKNV 251

Qy 237 VDPLTLKGLLELYERNR 255
 Db 252 INPFLTEGLLELYERNR 270

RESULT 6
 T36391
 hypothetical protein SCE94.31c - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
 C:Accession: T36391
 R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, April 1999
 A:Reference number: 221573
 A:Accession: T36391
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-265 <OLI>
 A:Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94.31c
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SCE94.31c
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 49.1%; Score 655; DB 2; Length 265;
 Best Local Similarity 51.0%; Pred. No. 9.4e-49;
 Matches 133; Conservative 46; Mismatches 74; Indels 8; Gaps 3;

Qy 1 LLLVIDGNTNTVLGVYHDKLEWHRIETSRHKTDEFGMLRSLFDHSGLMFEQ---- 56
 Db 1 MLLTIDVGNTHTVLGLFDGEDIVEHWRISTDSRRATDELAVLLQGLGMHPLGLDELGDG 60

Qy 57 IDGIISSVPPIMFALERMCTKYFHIEPOIVGPMKTGLNPKYDNPKEVGADRIVNAV 115
 Db 61 IDGIISSVPPIMFALERMCTKYFHIEPOIVGPMKTGLNPKYDNPKEVGADRIVNAV 120

Qy 116 AALHLYGNPLIVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETTR 175
 Db 121 AAVELYGCPAIVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETTR 180

Qy 176 PDNIIGNKTVSAMQSGILFGYGVQVEGIVKRMKWAQK----KODLKVIATGGLAPLIANESD 232
 Db 181 PRSVIGKNTVEAMQSGIVGYGAGQVGVNRMARELADDDPDDVTVIATGGLAPLMVLGESS 240

Qy 233 CIDIVDPFLTLKGLLELYERN 253
 Db 241 VIDHEPFLTLKGLLELYERN 261

RESULT 7
 B87489
 transcription activator, probable Baf family [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: B87489
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; MUID:21173698; PMID:11259647
 A:Accession: B87489
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-261 <STO>
 A:Cross-references: GB:AE005673; NID:gl3423392; PIDN:AAK23910.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC1935
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 43.2%; Score 577; DB 2; Length 261;
 Best Local Similarity 44.4%; Pred. No. 4.6e-42;
 Matches 114; Conservative 50; Mismatches 91; Indels 2; Gaps 1;

Qy 1 LLLVIDGNTNTVLGVYHDKLEWHRIETSRHKTDEFGMLRSLFDHSGLMFEQIDGI 60
 Db 2 MLLAIEQGNNTNTPAIDHGASWVAQWRSATSTRTADEYVYVWLSQLLSMOGLGPRADAV 61

Qy 61 IISVVPPIMFALERMCTKYFHIEPOIVGPMKTGLNPKYDNPKEVGADRIVNAVAIHL 120
 Db 62 IISVVPPIMFALERMCTKYFHIEPOIVGPMKTGLNPKYDNPKEVGADRIVNAVAIHL 121

Qy 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETTR--DN 178
 Db 122 YPGPLVVIDSGTATTFDVAADGAFFGIIAPGINLSMQALHHEAAAKLPRIAIQRPAGNR 181

Qy 179 IIGKNTVSAMQSGILFGYGVQVEGIVKRMKWAQKDLKVIATGGLAPLIANESDCIDI 238
 Db 182 IVGTDTVSAMQSGVFGVWGYISLIEGLVARIKAEGERPMTVIATGGLVASFEGATDSIDHFD 241

Qy 239 PFLTLKGLLELYERNR 255
 Db 242 SDLTIRGLLELYERNR 258

RESULT 8

H86937
 conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae
 C:Species: Mycobacterium leprae
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: H86937
 R: Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hickey, E.K.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, J.; Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 Nature 409, 1007-1011, 2001
 A: Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Massive gene decay in the leprosy bacillus.
 A: Reference number: A86909; PMID:21128732; PMID:11234002
 A: Accession: H86937
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-274 <STO>
 A: Cross-references: GB:AL450380; NID:gl3092576; PIDN:CAC29740.1; GSPDB:GN00147
 C: Gene: ML0232
 C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 38.9%; Score 519.5; DB 2; Length 274;
 Best Local Similarity 41.0%; Pred. No. 4.2e-37;
 Matches 109; Conservative 53; Mismatches 89; Indels 15; Gaps 5;

Qy 1 LLLVIDGNTNTVLGVY----HDGKLEYHWRIETSRHKTDEFGMILRSIFDHSGLMFEQ 56
 Db 1 MLLAIDVRNTHTVGLSSGMEHAKVVOQWRIRTESEVTADELALIDGLIGDSDS---ER 57

Qy 57 IDGIITSSVVPPIIMFALERMCTKYFHIPOI-VPGMKTGLNIDPNKPEVGADRIYNAV 115
 Db 58 LAGAAALSTVPSVLHEVRIMLDQWPSVPHVLEPGVIRGIPLLVDNPKPEVGADRIYNAV 117

Qy 116 AATHLYGNPLIVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETIR 175
 Db 118 AAYDRFRKAAIVVDFGSSICVDVSAKGEFLGGAIPAGVGVSSDAAASAAALRRVELAR 177

Qy 176 PDNIIGKNTVSAMQSGILFGYGVQVEGIVKRMKQAKQ---DI---KVIATGGLAPLIA 228
 Db 178 PRSVGKNTVECMQAGVFGAGLVGLVGRIRVEDVSGFSDVDHVAIVATGHTAPLLP 237

Qy 229 NESDCIDIVDFPFLTLKGLIYERNR 254
 Db 238 PELHTVDHYDRHLTLHGLRLVFERNR 263

RESULT 9
 A70955
 hypothetical protein Rv3600c - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: A70955
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, J.; Squares, S.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 Nature 393, 537-544, 1998
 A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; PMID:98295987; PMID:9634230
 A: Accession: A70955
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-272 <COL>
 A: Cross-references: GB:295557; GB:AL123456; NID:g3242276; PIDN:CAB08944.1; PID:g2113976
 A: Experimental source: strain H37Rv
 C: Gene: Rv3600c
 C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 37.0%; Score 494.5; DB 2; Length 272;
 Best Local Similarity 38.8%; Pred. No. 5.9e-35;
 Matches 102; Conservative 55; Mismatches 93; Indels 13; Gaps 4;

Qy 1 LLLVIDGNTNTVLGVY----HDGKLEYHWRIETSRHKTDEFGMILRSIFDHSGLMFEQ 56
 Db 1 MLLAIDVRNTHTVGLSSGMEHAKVVOQWRIRTESEVTADELALIDGLIGDSDS---ER 57

Qy 57 IDGIITSSVVPPIIMFALERMCTKYFHIPOI-VPGMKTGLNIDPNKPEVGADRIYNAV 115
 Db 58 LAGAAALSTVPSVLHEVRIMLDQWPSVPHVLEPGVIRGIPLLVDNPKPEVGADRIYNAV 117

Qy 116 AATHLYGNPLIVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETIR 175
 Db 118 AAYDRFRKAAIVVDFGSSICVDVSAKGEFLGGAIPAGVGVSSDAAASAAALRRVELAR 177

Qy 176 PDNIIGKNTVSAMQSGILFGYGVQVEGIVKRMKQAKQ---DI---KVIATGGLAPLIA 228
 Db 178 PRSVGKNTVECMQAGVFGAGLVGLVGRIRVEDVSGFSDVDHVAIVATGHTAPLLP 237

Qy 231 SDCIDIVDFPFLTLKGLIYERNR 253
 Db 238 LHTVDHYDQHLTLQGLRLVFERNR 260

RESULT 10
 E75516
 conserved hypothetical protein - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
 C:Accession: E75516
 R: White, C.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M. S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A: Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A: Reference number: A75250; PMID:20036896; PMID:10567266
 A: Accession: E75516
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-262 <WHI>
 A: Cross-references: GB:AE001905; GB:AE000513; NID:g6458144; PIDN:AAF10040.1; PID:g6458144
 A: Experimental source: strain R1
 C: Gene: DR0461
 C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 32.1%; Score 428; DB 2; Length 262;
 Best Local Similarity 37.4%; Pred. No. 2.9e-29;
 Matches 98; Conservative 51; Mismatches 99; Indels 14; Gaps 6;

Qy 2 LLLVIDGNTNTVLGVY--DGKLEYHWRIETSRHKTDEFGMILRSIFDHSGLMFEQIDGI 60
 Db 6 LLAVIDGNTTVGLADASGALHTWRTNREMLPDDLALQLHGLFTLAGAPIR--AA 63

Qy 61 IISVVVPII---MFALEMRCTKYFHIPOI-VPGMKTGLNIDPNKPEVGADRIYNAV 116
 Db 64 VLSVAPPVGVENVALAKR---HFMIDAFVSAENLPDVTVELDTPGSGADRLCNLFG 119

Qy 117 ALHLVG--NPLIVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIET 174
 Db 120 AEKYLGLGLDYAVVDFGTSTNFDVGRGRFRGLGGLIATGAQVSADALFAAALPRITIQ 179

Qy 175 RPDNIIGKNTVSAMQSGILFGYGVQVEGIVKRMKQAKQKQKQKQKQKQKQKQKQKQKQ 234
 Db 180 APETAIGKNTVHALQSLGVFGVAEMVDGLLRIRAEPLGAEAVAVATGFSRTVQCICQEI 239

Qy 235 DIVDPLTLKGLI-ELIYERNR 255
 Db 240 DYYDETTLRLGLVWASRSEV 261

RESULT 11
 D72320
 conserved hypothetical protein - Thermotoga maritima (strain MSB8)
 C:Species: Thermotoga maritima

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: D72320
R:Neilson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.
Nature 399, 323-329, 1999
A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: D72320
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-246 <ARN>
A:Cross-references: GB:AE001754; GB:AE000512; NID:94981417; PIDN:AD35964.1; PID:9498141
A:Experimental source: strain MSB8
C:Gene: TM0883
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 30.4%; Score 405.5; DB 2; Length 246;
Best Local Similarity 39.2%; Pred. No. 2.3e-27;
Matches 100; Conservative 42; Mismatches 88; Indels 25; Gaps 9;

Qy 1 LLLVLDVGNNTVLGVYHDKLEHYHRIETSRHKTDEFGMLRSLEPDH-----SGLMPEQ 56
Db 1 MLLVDVGNTHSVFSITDEGKTRFRWRSTGVQTEDE-----LFSHLPLLGDAWRE 53

Qy 57 IDGIISSVVPIMPALERMCTKYPHIEPQIVGPGMKTGL---NIKYNPKVGVADRIVN 113
Db 54 IKGIGVASVVTQNTVIERFSQKYPHISP--IWKAKNGCVKWNVK--NPSEVGADRIVN 109

Qy 114 AVAAHLYGNPLVVDFGTATTCYIDENKQYMGAIAPGITISTEALYSRAAKLPRIE 173
Db 110 VVAFVKEYGKNGIIMDMGTATTVDLV-VNGSYEGGAILPGPFMMVHSLFRGTAKLPLVEV 168

Qy 174 TRPNIIGKNTVSAMQSGILFGYQVEGIVKRMKQAKQDLKVIATGCGLAPLIANESDC 233
Db 169 KPAPFVVGKTEENIRLGVNNGSVYALEGIIIGRIK-EVYGLPVLVTGGQSKIV---KDM 224

Qy .234 I--DIVDFPLTKGL 246
Db 225 IKHEIFEDLLIKGV 239

RESULT 12.
D71326
conserved hypothetical protein TP0431 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 20-Jun-2000
C:Accession: D71326
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin they, L.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A>Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:96332770; PMID:9665876
A:Accession: D71326
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-273 <COL>
A:Cross-references: GB:AE001220; GB:AE000520; NID:93322705; PIDN:AAC65417.1; PID:9332271
A:Experimental source: strain Nichols
C:Gene: TP0431
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 24.5%; Score 327.5; DB 2; Length 273;
Best Local Similarity 29.9%; Pred. No. 1.3e-20;
Matches 75; Conservative 64; Mismatches 105; Indels 7; Gaps 4;

Qy 2 LLLVDVGNNTVLGVY--HDCK--LLEYHRIETSRHKTDEFGMLRSLEPDHSGLMPEQI 57
Db 1 MLLLDVGNHSHVFGIQGNGRVCVRELFRAPARKQDEYSLLIHALCERAGVGRASL 60

Qy 58 DGIHIISSVVPIMPALERMCTKYPHIEPQIVGPGMKTGLNIKYNP--KEVGADRIVNAV 115
Db 61 RDAFISSVWPVLTKTIADAVAQISGVQVVPVFGPWAYEHLVPRIPEVRAEITGDLVANAV 120

Qy 116 AAHLYGNPLVVDFGTATTCYIDENKQYMGAIAPGITISTEALYSRAAKLPRIETR 175
Db 121 AAYVHRSACVVVDCGALTFTAVDGTGLIQVAIAFGLRTAVQSLHTGTGAQLPLVPLAL 180

Qy 176 PDNIIGKNTVSAMQSGILFGYQVEGIVKRMKQAKQDLKVIATGCGLAPLIANESDCID 235
Db 181 PDSVLGKDTTHAVQAGVVRGTFLVIRAMIAQCQKELGCRCAAVITGGLSRFLFSSEVD-PP 239

Qy 236 IVPDFPLTKGL 246
Db 240 PIDAQLTSLGL 250

RESULT 13
F70165
conserved hypothetical protein BB0527 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: F70165
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A>Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: F70165
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-262 <KLE>
A:Cross-references: GB:AE001154; GB:AE000783; NID:92688431; PIDN:AAC66882.1; PID:9268843
A:Experimental source: strain B31

Query Match 20.3%; Score 271.5; DB 2; Length 262;
Best Local Similarity 27.3%; Pred. No. 7.9e-16;
Matches 70; Conservative 57; Mismatches 108; Indels 21; Gaps 5;

Qy 3 LVTDVGNNTVLGVYHDKLEHYHRIETSRHKTDEFGMLRSLEPDHSGLMPEQIDGIII 62
Db 9 LIIDIGNTSIAFALFKDNQVNLFIKMTNLMRLRYDEVYSPFEENFDEN-----VNKVI 62

Qy 63 SSVVPPIMPALERMCTKYPHIEPQIVGPGMKTGLNIKVD---NPKE-----VGADRIVNA 114
Db 63 SSVVPIINETFKNVIFSFYKIKPLFI-----GPDLYDLTFNPYKSDKFLGSDVFNAL 116

Qy 115 VAAHLYGNPLVVDFGTATTCYIDENKQYMGAIAPGITISTEALYSRAAKLPRIE 173
Db 117 VAALENYSFENLVVDLTGTACTIFAVSRQDGLGGLINSGLINFNLSLDNAYLIKKFPI 176

Qy 174 TRPNIIGKNTVSAMQSGILFGYQVEGIVKRMKQAKQDLKVIATGCGLAPLIANESDC 233
Db 177 STPNLLERTTSGVNSGLFYQYKLYEGVVRDIKQYKKKFNLIITGKNADLLSLIEI 236

Qy 234 IDVDPPLTKGLELI 249
Db 237 EFIFNIHLTVEGVAIL 252

RESULT 14
AI2292
hypothetical protein alt3896 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A>Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AI2292
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: A12292
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-276 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA075595.1; PID:gl7133030; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3896

Query Match 12.7%; Score 170; DB 2; Length 276;
Best Local Similarity 24.2%; Pred. No. 4.4e-07;
Matches 66; Conservative 50; Mismatches 115; Indels 42; Gaps 10;
QY 3 LVIDVGNNTVLGVYHDGKLEIYHRIE-----TSRHKTEDEFGMILRSLEFDHSGLM 53
DB 15 LALEIGNSRHLWALFNGESLEFTWDTYLPESVIOQLNGGETKLEVGSEKEIF-----FT 70
QY 54 FEQID-----GIIISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNICYDNPKVKG 107
DB 71 FFPPLPAPCPPLPLFIASVVPQOTVWE-----NYLNVRVITLD---QIPLNNIYPT---LG 120
QY 108 ADRIVNAVAHLYGNPLIIVDFGTATTCYIDENKOYMGGAIPAGITISTEALYSRAAK 167
DB 121 IDRALALWAGAGMSWGPFLVIDAGTALTFTADGCKNLVGGAILPGVGLQFASLGQQTGQ 180
QY 168 LPRIEITRPDNI---IGKNTVSAMQSGILFGYGVQVEGIVKRMKQAK-QDLKVIATGGL 223
DB 181 LPQVEMEAIKSLPPRFALNTTEAIQSGVIYTLIAGMRDFTF--EWLSLPDPDGKVAIKGGD 238
QY 224 APLIANESDCID-----IVDPFLTKGLELI 249
DB 239 RILLNLYQLALYDPLAARLIVEENLFWGMQTI 271

RESULT 15

S75559
hypothetical protein slr0812 - *Synechocystis* sp. (strain PCC 6803)
C;Species: *Synechocystis* sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S75559
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S75559
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-257 <KAN>
A;Cross-references: EMBL:D90911; GB:AB001339; NID:gl653083; PIDN:BA018120.1; PID:dl01885
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Start codon: GTG

Query Match 12.2%; Score 163; DB 2; Length 257;
Best Local Similarity 28.0%; Pred. No. 1.6e-06;
Matches 58; Conservative 42; Mismatches 79; Indels 28; Gaps 10;
QY 60 IITSSVVPPIPMFALERMCTKYFHI-EPOIVGPGMKTGLNICYDN-PKEVGADRIVNAVA 117
DB 64 LMLASVPE-----QTEVWRVYQFKIL-----TLKNLPLVNLVPSFGIDRALAGLT 110
QY 118 IHLVGNPLIIVDFGTATTCYIDENKOYMGGAIPAGITISTEALYSRAAKLPRIETR-- 175
DB 111 GLTYGPPCLAVDGTALTITGFPDQDKLVGGLATLPGQLQATLGDRLAALPKLEMDQLT 170
QY 176 --PDNIIGKNTVSAMQSGILFGYGVQVEGIVKRMKQAK-QDLKVIATGQ----LAPLIA 228
DB 171 ELDPDR-WALDTPSAIFSGVVYVGLGALQSYLQ--DMQKLPFGAAMVITGGDKILHGLK 227

QY 229 NESDCIDIV-DPFLTKGLELIYERNR 254
DB 228 EHSPLNSVANDDNLLIFLGMAAIHHGDR 254
Search completed: December 18, 2003, 13:03:35
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2003, 12:59:29 ; Search time 40 Seconds
(without alignments)
1664.439 Million cell updates/sec

Title: US-09-813-453B-2
Perfect score: 1335
Sequence: 1 LLLVDVGNNTVLGVYHDG.....PFLTKGLELIYERNRVGSV 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23.:

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1036	77.6	258	2 Q9F985	Q9F985 bacillus st
2	987	73.9	254	16 Q9KGH5	Q9KGH5 bacillus ha
3	853.5	63.9	255	16 Q8U15	Q8U15 oceanobacil
4	852	63.8	259	16 Q8YAC5	Q8YAC5 listeria mo
5	847	63.4	259	16 Q92F54	Q92F54 listeria in
6	846	63.4	259	2 Q8KU01	Q8KU01 listeria mo
7	775	58.1	273	16 Q97EB4	Q97EB4 clostridium
8	773	57.9	255	16 Q8R7M2	Q8R7M2 thermoaer
9	737	55.2	259	16 Q8XHL5	Q8XHL5 clostridium
10	655	49.1	265	16 Q9X8N6	Q9X8N6 streptomyce
11	577	43.2	261	16 Q9AGZ1	Q9AGZ1 caulobacter
12	539.5	40.4	256	16 Q8G5B8	Q8G5B8 bifidobacte
13	519.5	38.9	274	16 Q9CDS6	Q9CDS6 mycobacteri
14	503.5	37.7	257	16 Q8F7V9	Q8F7V9 leptospira
15	494.5	37.0	272	16 Q06282	Q06282 mycobacteri
16	458	34.3	256	16 Q8RFE4	Q8RFE4 fusobacteri

17	428	32.1	262	16 Q9RX54	Q9RX54 deinococcus
18	405.5	30.4	246	16 Q9WZV5	Q9WZV5 thermotoga
19	338.5	25.4	212	2 Q32514	Q32514 desulfovibr
20	327.5	24.5	273	16 Q83446	Q83446 treponema p
21	319	23.9	270	16 Q8KCK7	Q8KCK7 chlorobium
22	271.5	20.3	262	16 Q51477	Q51477 borrelia bu
23	185.5	13.9	255	16 Q8EUB0	Q8EUB0 mycoplasma
24	171	12.8	295	16 Q8Y2M4	Q8Y2M4 ralstonia s
25	170	12.7	276	16 Q8YOD7	Q8YOD7 anabaena sp
26	166	12.4	56	2 P94305	P94305 bacillus ps
27	166	12.4	254	16 Q8DUS3	Q8DUS3 synechococc
28	163	12.2	257	16 P74045	P74045 synechocyst
29	154.5	11.6	229	16 Q67753	Q67753 aquifex aeo
30	150	11.2	592	16 Q9JXF1	Q9JXF1 neisseria m
31	150	11.2	592	16 Q9JW17	Q9JW17 neisseria m
32	134.5	10.1	242	16 Q9PCI4	Q9PCI4 xylella fas
33	133	10.0	248	16 Q9HWC1	Q9HWC1 pseudomonas
34	119.5	9.0	242	16 Q8PFG5	Q8PFG5 xanthomonas
35	117	8.8	224	16 Q98Q93	Q98Q93 mycoplasma
36	116.5	8.7	223	16 Q9ZKY6	Q9ZKY6 helicobacte
37	116.5	8.7	242	16 Q8P3Y1	Q8P3Y1 xanthomonas
38	109.5	8.2	223	16 Q25533	Q25533 helicobacte
39	109	8.2	209	16 Q9PIA9	Q9PIA9 campylobact
40	104.5	7.8	597	17 Q30225	Q30225 archaeoglob
41	100	7.5	318	10 Q9LYV3	Q9LYV3 arabidopsis
42	95	7.1	520	16 Q8U6P6	Q8U6P6 agrobacteri
43	94.5	7.1	858	12 Q9QTD8	Q9QTD8 marek's dis
44	93.5	7.0	895	17 Q972N0	Q972N0 sulfobobus
45	93	7.0	676	12 Q8JTV2	Q8JTV2 lumpy skin

ALIGNMENTS

RESULT 1

Q9F985 PRELIMINARY; PRT; 258 AA.
 ID Q9F985
 AC Q9F985;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Putative 32 kDa replication protein.
 OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillales; Geobacillus.
 OX NCBI_TaxID=1422;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=V;
 RA Vasquez C., Pichuanes S., Saavedra C.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF198621; AAG28531.1; -
 DR InterPro; IPR004619; Baf.
 DR Pfam; PF03309; Bvg_acc_factor; 1.
 DR TIGRFAMs; TIGR00671; Baf; 1.
 SQ SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;

Query Match	77.6%	Score	1036;	DB	258;
Best Local Similarity	77.9%	Pred. No.	1.3e-85;		
Matches	197;	Conservative	31;	Mismatches	25;
		Indels	0;	Gaps	0;
Qy	1	LLLVDVGNNTVLGVYHDGKLEHYHRIETSRHKTEDEFGMILRSLPDHSGLMFEQIDGI	60		
Db	1	MIFVLDVGNNTVLGVYDGDGLKHHWRIETSRGTEDEYGMTIKALLNHVGLQSDIDGI	60		
Qy	61	IISGVPPPIFALERMCTKYFHIEPQIVGPMKTGLN KYDNKPKEVGADRVNAAVIAHL	120		
Db	61	IISGVPPPIFALERMCTKYFHIEPQIVGPMKTGLN KYDNKPKEVGADRVNAAVIAHL	120		
Qy	121	YGNPLIIVDGTATTTCYIDENKQYMGATAPGTITSTEALYSAAKLPIRIETRPDNI	180		
Db	121	YGSPLIIVDGTATTTCYIDENKQYMGATAPGTITSTEALYSAAKLPIRIETRPDNI	180		
Qy	181	GKNTVSAMQSGILFGYVQVEGIVKRMKQAKDLKVIATGTGLAPLIANESDCIDIVDPF	240		

Q97EB4 PRELIMINARY; PRT; 273 AA.
ID Q97EB4
AC Q97EB4;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Predicted transcriptional regulator, homolog of Bvg accessory

```
DE factor.
GN CAC3200.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RA MEDLINE=21359325; PubMed=11466286;
RX Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RA "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007815; AAK81136.1; -.
DR InterPro; IPR004619; Baf.
DR PFam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; 1.
KW Complete proteome.
SQ SEQUENCE 273 AA; 30331 MW; DE7B4D5923E72842 CRC64;

Query Match 58.1%; Score 775; DB 16; Length 273;
Best Local Similarity 56.4%; Pred. No. 5.8e-62;
Matches 146; Conservative 55; Mismatches 54; Indels 4; Gaps 1;

QY : 1 LLLVLDVGNNTVGLVYHDKLEYHWRISRTSHKTEDEFGMLRSLFDHSGLMFEQIDGI 60
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
12 VIILVDGNTNIVLGIYNDTKLTAENRLSDVLRSADEYGIQVWNLFDQDKLPTLVEGV 71

QY 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIHL 120
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
72 IISVVPPIMYSLHMRKYFKINPLVVGPGIKTGINKYDNPKEVGADRIVNAVAHEI 131

QY 121 YGNPLVVDGTTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETTRPDNI 180
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
132 YKSLIIDGTATTCVAVRENGDYLGGAICPGISSEALFEKAALPRVELIKPAYAI 191

QY 181 GKNTVSAMQSGILFGYGVQVEGIVKRMKQAKQDLKVIATGGLAPLIANESDCIDI 236
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
192 CKNTISSIQSGIVGIVGQRYIVRMKELQEGEKEPLVATGGLAKLISEAKNDV 251

QY 237 VDPFLTKGLELIYERNV 255
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
252 INPFLTLEGLRIIYERNV 270

RESULT 8
Q8R7M2 PRELIMINARY; PRT; 255 AA.
AC Q8R7M2;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE Putative transcriptional regulator, homologs of Bvg accessory
DE factor.
GN TTE2381.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=2192816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
```

```
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013180; AAM25520.1; -.
DR InterPro; IPR004619; Baf.
DR PFam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 255 AA; 27816 MW; C3C6205CB8CA6ED CRC64;

Query Match 57.9%; Score 773; DB 16; Length 255;
Best Local Similarity 57.1%; Pred. No. 8.1e-62;
Matches 145; Conservative 48; Mismatches 61; Indels 0; Gaps 0;

QY 1 LLLVLDVGNNTVGLVYHDKLEYHWRISRTSHKTEDEFGMLRSLFDHSGLMFEQIDGI 60
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1 MLLAFDVGNTNIVMGVFKGKLLHSFRISTDKNKTYDEYGLVNLQIYNGISLTIDDV 60

QY 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIHL 120
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 IISVVPPIMNTLQVMSLKYFKTKPIVVGPGIKTGINKYDNPKEVGADRIVNAVAAYEL 120

QY 121 YGNPLVVDGTTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETTRPDNI 180
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
121 YGPPVIVDFTATTCALISERGEYLGGLIAGELMISADALFORTAKLPKIDUTKPTVI 180

QY 181 GKNTVSAMQSGILFGYGVQVEGIVKRMKQAKQDLKVIATGGLAPLIANESDCIDI 240
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
181 NNTVASMQSGIYGHVGMVDYIVTRMGEPAPSAYVATGTFANMAIESKTIIDTVNEM 240

QY 241 LTLKGLELIYERNR 254
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
241 LTLEGLRIIYERNK 254

RESULT 9
Q8XHL5 PRELIMINARY; PRT; 259 AA.
AC Q8XHL5;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein CPE2468.
GN CPE2468.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP001394; BAB82174.1; -.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR000515; BPD_transp.
DR PFam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28819 MW; OD5FEA3B7A145E10 CRC64;

Query Match 55.2%; Score 737; DB 16; Length 259;
Best Local Similarity 56.2%; Pred. No. 1.5e-58;
Matches 145; Conservative 50; Mismatches 59; Indels 4; Gaps 1;

QY 1 LLLVLDVGNNTVGLVYHDKLEYHWRISRTSHKTEDEFGMLRSLFDHSGLMFEQIDGI 60
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1 MILLIDVGNNTIVLGHIDNEKYIASWRISTDSKTSDEYSIQVWLQFNQAKLPEDVEGI 60

QY 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIHL 120
```

```

Db 61 IISVVPPIMSLNMRKCKEPIVVGPGIKTGINKYDNPKEVGADRIVNAFAFEK 120
Qy 121 YGNPLIVVDFTATTTCYIDENKQYMGGAIAFGITISTEALYSRAAKLPRIETRPDNI 180
Db 121 HKKPMIIDFTATTTCYIDENKQYMGGAIAFGITISTEALYSRAAKLPRIETRPDNI 180
Qy 181 GNTVSAMOSGILFGYGVQVEGIVKRMKQWQ---AKQDLKVIATGGLAPLIANSDCIDI 236
Db 181 CKNTVTSMAQGIIVGKVEYIVKRMKMKEMMDLGEKEPFVLATGGLAKLVYSETDVIDE 240
Qy 237 VDPELTTLKGLBLIYERNR 254
Db 241 VDRKLTLEGLKILYEKNK 258

RESULT 10
Q9X8N6 PRELIMINARY; PRT; 265 AA.
AC Q9X8N6;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein SC03380.
GN SC03380 OR SCE94.31C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RA "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
[4]
RN SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=1200953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris H.M., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939116; CAB40880.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc.factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 49.1%; Score 655; DB 16; Length 265;
Best Local Similarity 51.0%; Pred. NO. 4.1e-51;

```

```

Matches 133; Conservative 46; Mismatches 74; Indels 8; Gaps 3;
Qy 1 LLLVIDVNTNTVLGVYHDKGLYHWRITSRHKTEDEFGMILRSFLPDHSGLMFEQ--- 56
Db 1 MLLTIDVGNTHVTLGLFDGEDIVHWRISTDSRTADELAVLLOGLMGMPHLLGDELGDG 60
Qy 57 IDGIISVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLINKYDNPKEVGADRIVNAV 115
Db 61 IDGIAICATVPSVLHRELVRTRYGVDPVAVLVEPVKTVGPIILTDHPKEVGADRIINAV 120
Qy 116 AATHLYGNPLIVVDFTATTTCYIDENKQYMGGAIAFGITISTEALYSRAAKLPRIETR 175
Db 121 AAVELYGGAIVVDFTATTTCYIDENKQYMGGAIAFGITISTEALYSRAAKLPRIETR 180
Qy 176 PDNIIGNKTVSAMOSGILFGYGVQVEGIVKRMKQWQ---KQDLKVIATGGLAPLIANS 232
Db 181 PRSVIGNKTVSAMOSGILFGYGVQVEGIVKRMKQWQ---KQDLKVIATGGLAPLIANS 240
Qy 233 CIDIVDPFLTLKGLBLIYERNR 253
Db 241 VIDEHEPWLTLGLRLVYERN 261

RESULT 11
Q9AGZ1 PRELIMINARY; PRT; 261 AA.
AC Q9AGZ1;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Transcriptional activator, putative, Baf family.
GN CC1935.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocak I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005867; AAK23910.1; -.
DR TIGR; CC1935;
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc.factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR Complete proteome.
SQ SEQUENCE 261 AA; 27965 MW; C19B60D7B0714EF5 CRC64;

Query Match 43.2%; Score 577; DB 16; Length 261;
Best Local Similarity 44.4%; Pred. NO. 4.6e-44;
Matches 114; Conservative 50; Mismatches 91; Indels 2; Gaps 1;
Qy 1 LLLVIDVNTNTVLGVYHDKGLYHWRITSRHKTEDEFGMILRSFLPDHSGLMFEQIDGI 60
Db 2 MLLAIEQNTNTVFAHIDGASVWAQWRSATSTRTADEYVVLSQLSMOGLGPRADAV 61
Qy 61 IISVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLINKYDNPKEVGADRIVNAVAIHL 120
Db 62 IISVVPPQSIFNLNLSRRYFNVPLVIGENAKLIGIDVRIEKPSEAGADRLVNAIGAMV 121
Qy 121 YGNPLIVVDFTATTTCYIDENKQYMGGAIAFGITISTEALYSRAAKLPRIETRP--DN 178
Db 122 YPGPLVVIDSGTATTDFIVAADGAFFGGIITAPGINSQALHEAAAKLPRIATQRPAGNR 181

```

```
QY 179 IIGKNTVSAMQSGILFGYGVQVEGIVKRMKQAKDLKVIATGGLAPLIANESDCIDIVD 238
Db 182 IVGTDIVSAMQSGVFGYGLISLIGLVARIKAERGEPMTVIATGCVASLFGEGATDSIDHFD 241

QY 239 PFLTLKGLLELIYERNR 255
Db 242 SDLTIRGLLEIYERNR 258

RESULT 12
Q8G558
ID Q8G558 PRELIMINARY; PRT; 256 AA.
AC Q8G558;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN Bll162.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karmirantzou M., Snel B., Villanova D., Berger B.,
RA Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
RA Pridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR EMBL; AB014740; AAN24967.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 256 AA; 27667 MW; 5C8BDE3AD6A3B5E CRC64;

Query Match 40.4%; Score 539.5; DB 16; Length 256;
Best Local Similarity 39.8%; Pred. No. 1.1e-40;
Matches 101; Conservative 53; Mismatches 99; Indels 1; Gaps 1;

QY 1 LLLVIDVGNNTVLGVYHDGKLEHYHWRITSRHKTEDEFGMLRSLFDHSGLMFEQIDGI 60
Db 1 MLVAVDIGNTNIVLGLDGLDAIAGTYRITTKANHTSDEYGLFTEFLRMSGFQPSVDVDD 60

QY 61 IISVVPPINFALERMCTKYFHTEPQIVGGMKTGLNPKYDNPKEVGADRIVNAVAIHL 120
Db 61 IICSVVPKVMHSFRSSIVKFLDIDPMVVGEGIKTMNVRVDDPKSLGADILADACAGAYE 120

QY 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETITRPDNI 180
Db 121 YGGPVLVADFGTATTTHVSDKGVDSGVITTGIRAGAAALWGDTAQLPEVEITRPDTIL 180

QY 181 GKNTVSAMQSGILFGYGVQVEGIVKRMKQW-AKQDLKVIATGGLAPLIANESDCIDIVDP 239
Db 181 GTNTKTCMQAGLYTFLGGVVERTIRQFRRLGEGDFKVIITGGLGRVFENDTELIDVYDP 240

QY 240 FLTLKGLLELIYERN 253
Db 241 DLIFKGMHIYSRN 254

RESULT 13
Q9CD56
ID Q9CD56 PRELIMINARY; PRT; 274 AA.
AC Q9CD56;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein ML0232.
GN ML0232.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
```

```
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornby T., Jagsels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrall B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
DR EMBL; AL583917; CAC29740.1; -.
DR Leproma; ML0232; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg acc. factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 274 AA; 29421 MW; 1C2E735BDEC8765 CRC64;

Query Match 38.9%; Score 519.5; DB 16; Length 274;
Best Local Similarity 41.0%; Pred. No. 8e-39;
Matches 109; Conservative 53; Mismatches 89; Indels 15; Gaps 5;

QY 1 LLLVIDVGNNTVLGVY---HDGKLEHYHWRITSRHKTEDEFGMLRSLFDHSGLMFQ 56
Db 1 MLLAIDVRNTHVTVGLLGSKEHAKVQVQWRITSEVTADELALIIDGLIGDDS---ER 57

QY 57 IDGIILSISSVPPINFALERMCTKYFHTEPQIVGGMKTGLNPKYDNPKEVGADRIVNAV 115
Db 58 LAGAAALSTVPSVLHEVRIMLDQYWPSPVPHVLIIEFGVRTGIPLLDNPKEVGADRIWCL 117

QY 116 AAHLVGNPLIVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETR 175
Db 118 AAFHFGQAAIVDFGSSICVDVSAKGEFLGATAPGVQVSSDAAAASAAALRRVELAR 177

QY 176 PNIIIGKNTVSAMQSGILFGYGVQVEGIVKRMKQAKQ---DL---KVIATGGLAPLIA 238
Db 178 PRSVVGKNTVECMQAGVYVFGAGLVGVRMRQVVEEFGDLGNRVAVVATGHTAPLL 237

QY 229 NESDCIDIVDFPLTLKGLLEIYERNR 254
Db 238 PELHTVVDHYDRHLTLHGLRLVFERNR 263

RESULT 14
Q8F7V9
ID Q8F7V9 PRELIMINARY; PRT; 257 AA.
AC Q8F7V9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Predicted transcriptional regulator.
GN LA0833.
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RA Ren S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB011268; AAN48032.1; -.
KW Complete proteome.
SQ SEQUENCE 257 AA; 28538 MW; 9AA0C72B32BA0B82 CRC64;

Query Match 37.7%; Score 503.5; DB 16; Length 257;
Best Local Similarity 41.0%; Pred. No. 2e-37;
Matches 105; Conservative 50; Mismatches 98; Indels 3; Gaps 3;
```

[illegible]

Search completed: December 18, 2003, 13:03:02
Job time : 43 secs

RESULT 15	006282	PRELIMINARY;	PRT;	272 AA.
ID	006282			
AC	006282;			
DT	01-JUL-1997 (TrEMBLrel. 04, Created)			
DT	01-JUL-1997 (TrEMBLrel. 04, Last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE	Hypothetical 29.3 kDa protein (Transcriptional activator, putative, Baf family).			
DE	RV3600C OR MTCV07H7B.22 OR MT3706.			
OS	Mycobacterium tuberculosis.			
OS	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
NCBI_TaxID=1773;				
ON	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=H37Rv;			
RC	MEDLINE=98295987; PubMed=9634230;			
EX				
RA	Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeter K., Gas S., Barry C.E. III, Tekala P.,			
RA	Badoeck K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davids R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornaby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skellern S., Squares S., Squares R.,			
RA	Sulton J.E., Taylor K., Whitehead S., Barrell B.G.;			
RT	Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence.;			
RL	Nature 393:537-544(1998).			
RP	[2]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RA	Fletschmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Petersson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,			
RA	Delcher A., Uterback T., Weidman J., Gill J., Mikula A.,			
RA	Bissh W.;			
RT	"whole genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains.;"			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
RL	EMBL; Z95557; CAB08944.1; -;			
DR	EMBL; AE007170; AAK48063.1; -;			
DR	TIGR; MT3706; -;			
DR	TubercuList; RV3600C; -;			
DR	InterPro; IPR004619; Baf.			
DR	Pfam; PF03309; Bvg_acc factor; 1.			
DR	TIGRFAMs; TIGR00671; baf; 1.			
KW	Hypothetical protein; Complete proteome.			
SQ	SEQUENCE 272 AA; 29304 MW; 5D70B6E6F0F09AC98 CRC64;			

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2003, 12:44:12 ; Search time 21 Seconds
(without alignments)
519.819 Million cell updates/sec

Title: US-09-813-453B-2

Perfect score: 1335

Sequence: 1 LLLVDVGNNTVTLGVYHDG.....PFLTLKGLIYERNRVGSV 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	181.5	13.6	249	US-09-328-352-4748	Sequence 4748, Ap
2	133	10.0	253	US-09-252-991A-21337	Sequence 21337, A
3	88	6.6	311	US-09-252-991A-23527	Sequence 23527, A
4	86	6.4	1011	US-09-252-991A-22505	Sequence 22505, A
5	83.5	6.3	207	US-08-549-515-9	Sequence 9, Appli
6	80	6.0	449	US-09-530-836-2	Sequence 2, Appli
7	80	6.0	502	US-09-530-836-2	Sequence 27619, A
8	79	5.9	252	US-09-328-352-5774	Sequence 5774, Ap
9	79	5.9	342	US-08-096-181A-12	Sequence 12, Appl
10	79	5.9	342	US-08-096-181A-14	Sequence 14, Appl
11	79	5.9	342	PCT-US94-08326-12	Sequence 12, Appl
12	79	5.9	342	PCT-US94-08326-14	Sequence 14, Appl
13	79	5.9	361	US-08-096-181A-8	Sequence 8, Appli
14	79	5.9	361	PCT-US94-08326-8	Sequence 8, Appli
15	79	5.9	363	US-08-096-181A-10	Sequence 10, Appl
16	79	5.9	363	PCT-US94-08326-10	Sequence 10, Appl
17	79	5.9	923	US-08-936-135-6	Sequence 6, Appli
18	78.5	5.9	357	US-09-120-051D-9	Sequence 9, Appli
19	78	5.8	1012	US-07-944-943-2	Sequence 2, Appli
20	78	5.8	1012	US-08-219-262B-2	Sequence 2, Appli
21	78	5.8	1012	US-09-031-655-2	Sequence 2, Appli
22	78	5.8	1213	US-08-937-102-2	Sequence 2, Appli
23	78	5.8	3898	US-08-876-991-2	Sequence 2, Appli
24	78	5.8	3898	US-09-059-853-2	Sequence 2, Appli
25	78	5.8	3898	US-08-750-717-2	Sequence 2, Appli
26	77.5	5.8	326	US-09-328-352-5506	Sequence 5506, Ap
27	77.5	5.8	1107	US-09-358-383C-16	Sequence 16, Appl

28	77	5.8	627	4	US-09-071-035-452	Sequence 452, App
29	77	5.8	1313	4	US-09-071-035-450	Sequence 450, App
30	77	5.8	1313	4	US-09-071-035-454	Sequence 454, App
31	76	5.7	922	4	US-09-116-473-4	Sequence 4, Appli
32	75.5	5.7	443	4	US-09-594-193-9	Sequence 9, Appli
33	75.5	5.7	551	4	US-09-328-352-7176	Sequence 7176, Ap
34	75.5	5.7	1215	4	US-09-134-001C-5319	Sequence 5319, Ap
35	75	5.6	553	6	5310678-1	Patent No. 5310678
36	75	5.6	1012	1	US-08-219-262B-9	Sequence 9, Appli
37	75	5.6	1012	3	US-09-031-655-9	Sequence 9, Appli
38	74.5	5.6	246	4	US-09-634-238-224	Sequence 224, App
39	74.5	5.6	262	4	US-09-252-991A-21196	Sequence 21196, A
40	74.5	5.6	452	4	US-09-252-991A-31196	Sequence 31196, A
41	74.5	5.6	452	4	US-09-134-001C-3838	Sequence 3838, Ap
42	74	5.5	413	3	US-09-147-928-2	Sequence 2, Appli
43	73.5	5.5	416	4	US-09-328-352-6675	Sequence 6675, Ap
44	73.5	5.5	473	4	US-09-107-532A-5371	Sequence 5371, Ap
45	73.5	5.5	499	4	US-09-328-352-5089	Sequence 5089, Ap

ALIGNMENTS

RESULT 1

US-09-328-352-4748
; Sequence 4748, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328.352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4748
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4748

Query Match 13.6%; Score 181.5; DB 4; Length 249;

Best Local Similarity 27.6%; Pred. No. 4.8e-13;

Matches 79; Conservative 39; Mismatches 73; Indels 95; Gaps 18;

Qy	3	LVIDVGNNTVTLGVYHDGKLEYHWRITSR---HKTEDE-----FGMTLRSLFDHSG 51
Db	9	LWLDIGNT-----RLKY-WITENQOIIEHAAELHLOSPADLLGLQH--FKHQG 55
Qy	52	LMFEQIDGIITSSVVP-----PIMFALERMCTKYFHIEPQIVGPGMKTG 95
Db	56	L-----HRIGISSVLDTENNQRIQIILKWLIEIPVFA--KVHAEY-----AG 95
Qy	96	LNTRYDNPKEVGADRIYNVAAHLYGNPLIVDFGATTTCYID--ENKQYMGGAIPG 153
Db	96	LQCGYVPSQLGIDRWLQVLVAEKEN-YCIICGCTALT---IDLTGKHGLGGLYLR 151
Qy	154	ITISTEALY--SRAAKLPRIETFPDNI-IGKNTVSAMQSGILFGYGVQGVGIVKRMKW 210
Db	152	LYLQORDALIONTKGIKIPDSAF---DNINPGNNTVDVHHGILGLISTIESIMQ----- 203
Qy	211	AKQDKVIATGGLAPLIAN-----ESDCIDIVDPFLLTKGLE 247
Db	204	-QSPKLLLTGSDAPLFAKFLQKYQPTVEID-----LLLLKGLQ 240

RESULT 2

US-09-252-991A-21337
; Sequence 21337, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 21337
 LENGTH: 253
 TYPE: PRP
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-21337

Query Match 10.0%; Score 133; DB 4; Length 253;
 Best Local Similarity 25.7%; Pred. No. 2.7e-07;
 Matches 46; Conservative 35; Mismatches 80; Indels 18; Gaps 5;
 QY 81 FHIEPQIVGPMK-TGLNIKYDNPKEVGADRIVNAVAATHLYGNPLIVVDGPTATTCYVI 139
 DB 77 FVSALVASSGKQAGVRNGYLDYQRLGDRWALVAHHAKKACLVLDLGTAVTSDLV 136
 QY 140 DENKQYMGGAIPGTTISTEALYSRAAKLPRIEITPDNIIGNVTVSAMQSG-----IL 193
 DB 137 AADGVHLGGVYICPGMTLMRSQLRTHTRI-----RYDDAEARALASLQPGQATAEAVE 190
 QY 194 FGTVGQVEGIVKRMKQAKO-----DLKVIATGGLAPLANESDCIDIVDPFLTLKGLLEL 248
 DB 191 RGCLLMRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIM-PDLVFVGLAL 248

RESULT 3

US-09-252-991A-23527
 Sequence 23527, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 23527
 LENGTH: 311
 TYPE: PRP
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-23527

Query Match 6.6%; Score 88; DB 4; Length 311;
 Best Local Similarity 22.7%; Pred. No. 0.079;
 Matches 60; Conservative 45; Mismatches 79; Indels 80; Gaps 15;
 QY 5 IDVGNNTVLGVYHDGKLEHYHRIETSRHKTEDEFGMLRLSFLDHSGLMFEQIDGIIIS 64
 DB 66 LDVDNGRVKGVKPFENIRDAGDPVEIARR-----YDEQGA--DEITFLDITA 110
 QY 65 VV---PPIMALERMCTKYHIEPQIVGPMKTLGAIKYNPNKEVGADRIVNAVAATHLY 121
 DB 111 SVDRGDTTLTHTVERMASQVF--IPLTVGGVRSVQDIR--NLLNAGADKVSINTAAVF-- 164
 QY 122 GNPLIVVD---FGTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIE----- 173
 DB 165 -NPEFGEAADRGSCQIVVAIDAKK---EVSAPG-----EAPWEIFTHG. 205
 QY 174 -----TRPDNI-----GKNTVSAM--QSGILFGY-VGQVEGIVKRMKQAKQDLKV 217
 DB 206 GRKPTGLDAVLWAKKMDLGAIBILLTSMDDQGVKSGYDLGVTRAISEAV-----NVPV 259

QY 218 IATGGL-----APLIANESDCI 234
 DB 260 IASGGVGNLEHLAAGILEGKADAV 283
 RESULT 4
 US-09-252-991A-22505
 Sequence 22505, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 22505
 LENGTH: 1011
 TYPE: PRP
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-22505

Query Match 6.4%; Score 86; DB 4; Length 1011;
 Best Local Similarity 21.8%; Pred. No. 0.87;
 Matches 52; Conservative 30; Mismatches 79; Indels 78; Gaps 11;
 QY 47 FDHSGLMFEQIDGIIISVVPPIMFALE--RMCTKYEHIEPQIVGPMKTLGAIKYNPNK 104
 DB 676 FLEEGLVGGVGLAGASLVPADVGLGFEALGEOGFIILRRQVVDHGLRTG-----PE 728
 QY 105 EVGADRIVNAVAATHLYGNPLIVVDG-----TATTYCYIDENKQYMGGAIPGTTIS 157
 DB 729 RYGRDAGSGDGFVH-----EVEQDFGLQADTLNFAHCLPHS-----A 768
 QY 158 TVALYSRAAKLPRIEITPDNIIGNVTVSAMQSGILFGVQVEGIVKRMK----- 209
 DB 769 LDQLWTRAA-----RP-----CSPSRGILHDAAGKISGYASRYDWMTLHGHVG 811
 QY 210 --QAKQ-----DLKVIATGGLAPLIANESDCIDIVDPFLTLKGLLELYERNRVGSV 258
 DB 812 QAQCSQKNCITLFOINKLTVHDLRAQLRRGVAGC-----FTALRA-PCLYSRNRPRSL 863

RESULT 5

US-08-549-515-9
 Sequence 9, Application US/08549515
 Patent No. 6054123
 GENERAL INFORMATION:
 APPLICANT: Loosmore, Sheena M
 APPLICANT: Klein, Michel H
 TITLE OF INVENTION: Haemophilus Influenzae
 TITLE OF INVENTION: Dimethylsulphoxide Reductase Enzyme
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Sim & McBurney
 STREET: Suite 701, 330 University Avenue
 CITY: Toronto
 STATE: Ontario
 COUNTRY: Canada
 ZIP: M5G 1R7
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/549,515

```

Query Match          5.9%; Score 79; DB 4; Length 252;
Best Local Similarity 21.1%; Pred. No. 0.65;
Matches 55; Conservative 49; Mismatches 83; Indels 74; Gaps 16;

Qy 7 VGNNTVTVGYHDGKLEY---HWR-----IETSRHKTEDE-----F 39
   ||| : : : : : : : : : : : : : : : : : : : : : :
Ddb 17 IDSIRTVL---FFNGEINFKKKWSKNITGLEISNEMTQSENGRLIQYVEITLDSNKQF 74
   ||| : : : : : : : : : : : : : : : : : : : : : :
Qy 40 GMIL---RSLFDHSGLMGFEOIDGI-ITSSVVPPIFMALERCTKYFHIEPQIVGPMKGTG 95

```

Db 75 NLVYLKQSLIDLQ-LVFERDENFYTFNEIKVEDFPYKISVFPQLNEKLIIRGNVVE 133
QY 96 LNIKYDNPKEVGADRIVNAVAIHLVGNPLVVDFGTATTCYVIDENK-----QYMG-- 148
Db 134 LSIPIVDNEK-IGDILRSNVSVLNNMQEDLEEISYRTNKS-FIDNIKINQVQVYSGOK 191
QY 149 ---AIAPGITISTEALYSRAAKLPRIETIRPONI---ICKNTVSAMQSGILF-GYVGQVE 201
Db 192 MSUVDPNIGI-----PKRVOK--NILMNDVNTDASHRSELDLKFIPLLQ 237
QY 202 GIVKRMKQAKDLKVIATGG 222
Db 238 DSVK-----KLIRNGG 248
RESULT 9
US-08-096-181A-12
; Sequence 12, Application US/08096181A
; Patent No. 6153406
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Pullen, Jeffrey K.
; APPLICANT: Soper, Thomas S.
; APPLICANT: Liang, Shu-Mei
; TITLE OF INVENTION: A Method For The High Level Expression,
; Purification And Refolding Of The Outer Membrane Protein
; TITLE OF INVENTION: P2 From Haemophilus Influenzae Type B
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/096,181A
; FILING DATE: 23-Jul-1993
; CLASSIFICATION: 424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-096-181A-12
Query Match 5.9%; Score 79; DB 3; Length 342;
Best Local Similarity 20.1%; Pred. No. 1.1; Indels 62; Gaps 10;
Matches 43; Conservative 38; Mismatches 71; Indels 62; Gaps 10;
QY 28 IETSRHKT-----EDEFGLMILRSLF-----DHSGLMFEQIDGIIISVVPPIMFAL 73
Db 100 VKLGRAKTIADGITSADKEYGLVNNSDYPTSGNTVGYTFKIDGLVLA---NYLLAQ 156
QY 74 ERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAIHLVGNPLVVDFGTA 133
Db 157 KREGAKGNKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR----- 203
QY 134 TTICY--IDENKQYMGAIAP-----GITISTEALYSRAAKLP-----R 170
Db 204 TNYKNESDEHKQOLNGVLATLGYRPSDLGLLVSLDSGAKTKNYKIKHEKRYFVSPGQ 263
QY 171 IEITRPDNIIG-----KNTV-----SAMQSGILFG 195
Db 264 YELMEDTNYVGNFKYERTSVDOGEKTRQAVLFG 297
RESULT 11
PCT-US94-08326-12
; Sequence 12, Application PC/TUS9408326
; GENERAL INFORMATION:
; APPLICANT: North American Vaccine, Inc.
; APPLICANT: 12103 Indian Creek Court
; APPLICANT: Beltsville, MD 20705
; APPLICANT: Pullen, Jeffrey K.

Db 264 YELMEDTNYVGNFKYERTSVDOGEKTRQAVLFG 297
RESULT 10
US-08-096-181A-14
; Sequence 14, Application US/08096181A
; Patent No. 6153406
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Pullen, Jeffrey K.
; APPLICANT: Soper, Thomas S.
; APPLICANT: Liang, Shu-Mei
; TITLE OF INVENTION: A Method For The High Level Expression,
; Purification And Refolding Of The Outer Membrane Protein
; TITLE OF INVENTION: P2 From Haemophilus Influenzae Type B
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/096,181A
; FILING DATE: 23-Jul-1993
; CLASSIFICATION: 424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-096-181A-14
Query Match 5.9%; Score 79; DB 3; Length 342;
Best Local Similarity 20.1%; Pred. No. 1.1; Indels 62; Gaps 10;
Matches 43; Conservative 38; Mismatches 71; Indels 62; Gaps 10;
QY 28 IETSRHKT-----EDEFGLMILRSLF-----DHSGLMFEQIDGIIISVVPPIMFAL 73
Db 100 VKLGRAKTIADGITSADKEYGLVNNSDYPTSGNTVGYTFKIDGLVLA---NYLLAQ 156
QY 74 ERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAIHLVGNPLVVDFGTA 133
Db 157 KREGAKGNKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR----- 203
QY 134 TTICY--IDENKQYMGAIAP-----GITISTEALYSRAAKLP-----R 170
Db 204 TNYKNESDEHKQOLNGVLATLGYRPSDLGLLVSLDSGAKTKNYKIKHEKRYFVSPGQ 263
QY 171 IEITRPDNIIG-----KNTV-----SAMQSGILFG 195
Db 264 YELMEDTNYVGNFKYERTSVDOGEKTRQAVLFG 297
RESULT 11
PCT-US94-08326-12
; Sequence 12, Application PC/TUS9408326
; GENERAL INFORMATION:
; APPLICANT: North American Vaccine, Inc.
; APPLICANT: 12103 Indian Creek Court
; APPLICANT: Beltsville, MD 20705
; APPLICANT: Pullen, Jeffrey K.

```
APPLICANT: Soper, Thomas S.
APPLICANT: Liang, Shu-Mei
TITLE OF INVENTION: A Method For The High Level
TITLE OF INVENTION: Expression,
TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 14
TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/08326
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,181
FILING DATE: 23-JULY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REFERENCE/DOCKET NUMBER: 1438.001PC01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-08326-12

Query Match 5.9%; Score 79; DB 5; Length 342;
Best Local Similarity 20.1%; Pred. No. 1.1;
Matches 43; Conservative 38; Mismatches 71; Indels 62; Gaps 10;

Qy 28 IETSRHKT-----EDEFGMILRSLF-----DHSLMFEQIDGIISSVWPPIMFAL 73
Db 100 VLKGRAKTIADGITSADKEYGVLNNSDIPTSGNTVGYTFKGDGLVGA---NYLLAQ 156
Qy 74 ERMCTKYFHIEPQIVGPMKMTGLNIKYDNPKEVGADRIVNAVAIHLVGNPLIVVDFGTA 133
Db 157 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR----- 203
Qy 134 TTTCY--IDENKQVMGGAIAAP-----GITISTEALYSRAAKLP-----R 170
Db 204 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLVSLDSGYAKTKNYKIKHKRYFVSPGFQ 263
Qy 171 IEITRPNIIIG-----KNTV-----SAMQSGILFG 195
Db 264 YELMEDTNNVGNFKYERTSVDDQGEKTRQAVLFG 297

RESULT 12
PCT-US94-08326-14
; Sequence 14, Application PC/TUS9408326
; GENERAL INFORMATION:
; APPLICANT: North American Vaccine, Inc.
; APPLICANT: 12103 Indian Creek Court
; APPLICANT: Beltsville, MD 20705
; APPLICANT: Pullen, Jeffrey K.
; APPLICANT: Soper, Thomas S.
; APPLICANT: Liang, Shu-Mei
; TITLE OF INVENTION: A Method For The High Level
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane Protein
; TITLE OF INVENTION: Expression,
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane Protein
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
```

```
APPLICANT: Soper, Thomas S.
APPLICANT: Liang, Shu-Mei
TITLE OF INVENTION: A Method For The High Level
TITLE OF INVENTION: Expression,
TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 14
TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/08326
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,181
FILING DATE: 23-JULY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REFERENCE/DOCKET NUMBER: 1438.001PC01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-08326-14

Query Match 5.9%; Score 79; DB 5; Length 342;
Best Local Similarity 20.1%; Pred. No. 1.1;
Matches 43; Conservative 38; Mismatches 71; Indels 62; Gaps 10;

Qy 28 IETSRHKT-----EDEFGMILRSLF-----DHSLMFEQIDGIISSVWPPIMFAL 73
Db 100 VLKGRAKTIADGITSADKEYGVLNNSDIPTSGNTVGYTFKGDGLVGA---NYLLAQ 156
Qy 74 ERMCTKYFHIEPQIVGPMKMTGLNIKYDNPKEVGADRIVNAVAIHLVGNPLIVVDFGTA 133
Db 157 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR----- 203
Qy 134 TTTCY--IDENKQVMGGAIAAP-----GITISTEALYSRAAKLP-----R 170
Db 204 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLVSLDSGYAKTKNYKIKHKRYFVSPGFQ 263
Qy 171 IEITRPNIIIG-----KNTV-----SAMQSGILFG 195
Db 264 YELMEDTNNVGNFKYERTSVDDQGEKTRQAVLFG 297

RESULT 13
US-08-096-181A-8
; Sequence 8, Application US/08096181A
; Patent No. 6153406
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Pullen, Jeffrey K.
; APPLICANT: Soper, Thomas S.
; APPLICANT: Liang, Shu-Mei
; TITLE OF INVENTION: A Method For The High Level Expression,
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane Protein
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane Protein
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
```

ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,181A
FILING DATE: 23-Jul-1993
CLASSIFICATION: 424
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-096-181A-8

Query Match 5.9%; Score 79; DB 3; Length 361;
Best Local Similarity 20.1%; Pred. No. 1.2;
Matches 43; Conservative 38; Mismatches 71; Indels 62; Gaps 10;
QY 28 IETSRHKT-----EDEFGLMRLSLF-----DHSGLMFEQIDGIISSVVPPIPMFAL 73
DB 119 VKLGRKTIADGITSADKEYGVNNSDYPTSGNTVGTFTKIDGLVGA---NYLLAQ 175
QY 74 ERMCTKYPHIEPQIVGPGMKTLNPKYGVADRIVNAVAHIHLYGNPLIVDFGTA 133
DB 176 KREGAKGENKRPNDKAGEVRIG---EINNGIQVAKYDANDIVAKIAYGR----- 222
QY 134 TTYCY--IDENKQYMGATAP-----GITISTEALYSRAKLP-----R 170
DB 223 TNYKYNESDEHKQQLNGVLATLGYRPSDLGLVSLDSGYAKTKNYIKHKEKRYFVSPGFQ 282
QY 171 IEITRPDNIIG-----KNTV-----SAMQSGILFG 195
DB 283 YELMEDTNYGNFKYERTSVDOGEKTRQAVLFG 316

RESULT 14
PCT-US94-08326-8
Sequence 8, Application PC/TUS9408326
GENERAL INFORMATION:
APPLICANT: North American Vaccine, Inc.
APPLICANT: 12103 Indian Creek Court
APPLICANT: Beltsville, MD 20705
APPLICANT: Pullen, Jeffrey K.
APPLICANT: Soper, Thomas S.
APPLICANT: Liang, Shu-Mei
TITLE OF INVENTION: A Method For The High Level
TITLE OF INVENTION: Expression,
TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane
TITLE OF INVENTION: Protein
TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/08326
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,181
FILING DATE: 23-JULY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Emond, Robert W.
REFERENCE/DOCKET NUMBER: 1438.001PC01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-08326-8

Query Match 5.9%; Score 79; DB 5; Length 361;
Best Local Similarity 20.1%; Pred. No. 1.2;
Matches 43; Conservative 38; Mismatches 71; Indels 62; Gaps 10;
QY 28 IETSRHKT-----EDEFGLMRLSLF-----DHSGLMFEQIDGIISSVVPPIPMFAL 73
DB 119 VKLGRKTIADGITSADKEYGVNNSDYPTSGNTVGTFTKIDGLVGA---NYLLAQ 175
QY 74 ERMCTKYPHIEPQIVGPGMKTLNPKYGVADRIVNAVAHIHLYGNPLIVDFGTA 133
DB 176 KREGAKGENKRPNDKAGEVRIG---EINNGIQVAKYDANDIVAKIAYGR----- 222
QY 134 TTYCY--IDENKQYMGATAP-----GITISTEALYSRAKLP-----R 170
DB 223 TNYKYNESDEHKQQLNGVLATLGYRPSDLGLVSLDSGYAKTKNYIKHKEKRYFVSPGFQ 282
QY 171 IEITRPDNIIG-----KNTV-----SAMQSGILFG 195
DB 283 YELMEDTNYGNFKYERTSVDOGEKTRQAVLFG 316

RESULT 15
US-08-096-181A-10
Sequence 10, Application US/08096181A
Patent No. 6153406
GENERAL INFORMATION:
APPLICANT: Tai, Joseph Y.
APPLICANT: Pullen, Jeffrey K.
APPLICANT: Soper, Thomas S.
APPLICANT: Liang, Shu-Mei
TITLE OF INVENTION: A Method For The High Level Expression,
TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane
TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,181A
FILING DATE: 23-Jul-1993
CLASSIFICATION: 424

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 363 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-096-181A-10

Query Match 5.9%; Score 79; DB 3; Length 363;
Best Local Similarity 20.1%; Pred. No. 1.2;
Matches 43; Conservative 38; Mismatches 71; Indels 62; Gaps 10;
Qy 28 IETSRHKT-----EDEFQMILRSLF-----DHGSLMFEQIDGIIISVVPPIPFAL 73
Db 121 VKLGRAKTIADGITSADKEKGVGLNNSDYIPTSGNTVGYPFGIDGLVLGA---NYLLAQ 177
Qy 74 ERMCTKYPHIEPQIVGPMKTLNLIKYNPNKEVGADRIVNAVAATHLYGNPLIVVDFGTA 133
Db 178 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR----- 224
Qy 134 TTYCY--IDENKQYMGGAIP-----GITISTEALYSRAAKLP-----R 170
Db 225 TNYKYNESDEHKQQLNGVLATLGYRFDLGLLVSLDSCYAKTKNYKIKHKRYFVSPGFQ 284
Qy 171 IEITRPDNIIG-----KNTV---SAMQSGILFG 195
Db 285 YELMEDTNVGNFKYRTSVDOGEKTRQAVLFG 318

Search completed: December 18, 2003, 12:56:17
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2003, 12:56:23 ; Search time 17 Seconds
(without alignments)
713.699 Million cell updates/sec

Title: US-09-813-453B-2

Perfect score: 1335

Sequence: 1 LLLIVDVGNTNTVLGVYHDG.....PFLTLKGLIYERNRVGSV 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1114	83.4	233	1 YACB_BACSU	P37564 bacillus su
2	129.5	9.7	267	1 BAF_BORPE	Q45338 bordetella
3	90.5	6.8	4451	1 GRSE_BACBR	P14688 b gramicidin
4	89.5	6.7	357	1 BUK_THETN	Q8r832 thermoanaer
5	88.5	6.6	425	1 PURA_LEPIN	Q8f738 leptospira
6	88	6.6	255	1 HT61_PSEAE	Q9hu44 pseudomonas
7	87.5	6.6	424	1 SAHH_METKA	P38855 methanopyru
8	87	6.5	336	1 PYRD_SALTI	Q8z769 salmonella
9	87	6.5	336	1 PYRD_SALTY	P25468 salmonella
10	86.5	6.5	243	1 Y004_METJA	Q60315 methanococc
11	85	6.4	336	1 PYRD_ECOLI6	Q8fj91 escherichia
12	85	6.4	336	1 PYRD_ECOLI	P05021 escherichia
13	83.5	6.3	204	1 DMSB_ECOLI	P18776 escherichia
14	83	6.2	187	1 ATKC_RHILO	Q989x7 rhizobium l
15	83	6.2	208	1 P1MT_VIBPA	Q87lq6 vibrio para
16	83	6.2	300	1 NANK_HAEIN	P44541 haemophilus
17	83	6.2	425	1 GSA_CLOPE	Q9znc8 clostridium
18	83	6.2	470	1 CH60_TRIVA	Q95058 trichomonas
19	83	6.2	628	1 GATE_SULTO	Q97iw7 sulfolobus
20	82	6.1	305	1 GALE_METJA	Q57664 methanococc
21	82	6.1	3898	1 POLG_HCVB	P21530 hog cholera
22	81	6.1	359	1 OP21_HAEIN	P43839 haemophilus
23	81	6.1	449	1 MURD_STRAW	Q8nx35 staphylococ
24	81	6.1	1224	1 RPB2_YEAST	P08518 saccharomyc
25	80.5	6.0	208	1 P1MT_YERPE	Q8zbq0 yersinia pe
26	80.5	6.0	297	1 NANK_PAGMU	Q2ckb3 pasteurella
27	80.5	6.0	327	1 EBERG_ECOLI	P06846 escherichia
28	80.5	6.0	354	1 LUXA_VIBFI	P19907 vibrio fisc
29	80.5	6.0	410	1 RAA2_CHLRE	Q9emb4 chlamydomon
30	80	6.0	199	1 ATC2_ANASP	Q8ypf1 anabaena sp
31	80	6.0	449	1 MURD_STRAAM	Q33595 staphylococ
32	79.5	6.0	346	1 GCP_BACSU	O05518 bacillus su
33	79	5.9	304	1 HEM3_AQUAE	O66621 aquifex aeo

RESULT 1

ID	YACB_BACSU	STANDARD;	PRT;	233 AA.
AC	P37564;			
DT	01-OCT-1994 (Rel. 30, Created)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Hypothetical protein yacB.			
GN	YACB.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
ON	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=96051385; PubMed=7584024;			
RA	Ogasawara N., Nakai S., Yoshikawa H.;			
RT	"Systematic sequencing of the 180 kilobase region of the Bacillus			
RL	subtilis chromosome containing the replication origin.";			
RN	DNA Res. 1:1-14(1994).			
RP	[2]			
RC	SEQUENCE FROM N.A.			
RX	STRAIN=168;			
RA	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,			
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,			
RA	Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,			
RA	Guiseppi G., Guy B.J., Haga K., Haiesch J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,			
RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,			
RA	Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,			
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,			
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,			
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,			
RA	Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,			
RA	Presecan E., Pujic C., Purnelle B., Rapoport G., Rey M., Reynolds S.,			
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,			
RA	Sato T., Scanlan J., Schleich S., Schroeter R., Scoffone F.,			
RA	Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,			
RA	Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,			
RA	Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,			
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,			
RA	Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,			
RA	Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,			
RA	Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,			
RT	"The complete genome sequence of the Gram-positive bacterium Bacillus			
RT	subtilis.";			
CC	Nature 390:249-256(1997).			
CC	-!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.			

Q57651 methanococc
P20149 haemophilus
P97333 mus musculus
Q9aj64 rickettsia
Q91603 mus musculus
P32925 geotrichum
Q10334 schizosacch
Q8zmf9 salmonella
Q8ub9 agrobacteri
Q9wv27 mus musculu
Q8xhb3 clostridium
P33345 escherichia

ALIGNMENTS

```

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; D26185; BAA05305.1; -.
DR EMBL; Z99104; CAB11846.1; -.
DR PIR; S66100; S66100.
DR Subtilist; BG10133; yacB.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; Baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 233 AA; 26217 MW; AAE96E732C15DF44 CRC64;

Query Match      83.4%; Score 1114; DB 1; Length 233;
Best Local Similarity 99.5%; Pred. No. 2e-85; Indels 0; Gaps 0;
Matches 233; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLVIDVGNNTVGLGVYHDGKLEHYHRIETSRHKTDEFGMILSLFDHSGLMFEQIDGI 60
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MLLVIDVGNNTVGLGVYHDGKLEHYHRIETSRHKTDEFGMILSLFDHSGLMFEQIDGI 60

QY 61 IISVVPPIMFALERMCTKFFHIEPOIVGPMKTGLNPKYDNPKEVGADRIVNAVAIHL 120
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 IISVVPPIMFALERMCTKFFHIEPOIVGPMKTGLNPKYDNPKEVGADRIVNAVAIHL 120

QY 121 YGNPLIVDPGTATTTCYIDENQYMGGAAPGTTISTEALYSRAAKLPRIETTRPDNII 180
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 YGNPLIVDPGTATTTCYIDENQYMGGAAPGTTISTEALYSRAAKLPRIETTRPDNII 180

QY 181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWAQKOD 214
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWAQKOD 214

RESULT 2
BAF BORPE
ID _BAF BORPE STANDARD; PRT; 267 AA.
AC Q45338; Q45373;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Bvg accessory factor.
DE Bacteriophage T4.
DE Bordetella pertussis.
DE Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
DE Alcaligenaceae; Bordetella.
DE NCB1_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP504;
RX MEDLINE=95325323; PubMed=7601846;
RA Deshaizer D., Wood G.E., Friedman R.L.;
RT "Identification of a Bordetella pertussis regulatory factor required
RT for transcription of the pertussis toxin operon in Escherichia
RT coli."
RL J. Bacteriol. 177:3801-3807(1995).
RN [2]
RP SEQUENCE OF 1-38 FROM N.A.
RC STRAIN=BP504;
RA Wood G.E., Friedman R.L.;
RT "Identification of a bira homolog in Bordetella pertussis."
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 239-267 FROM N.A.
RC STRAIN=BP536;
RX MEDLINE=96419162; PubMed=8821935;
RA Allen A.G., Maskell D.J.;

```

```

RT "The identification, cloning and mutagenesis of a genetic locus
RT required for lipopolysaccharide biosynthesis in Bordetella
RT pertussis."
RL Mol. Microbiol. 19:37-52(1996).
CC -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
CC A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
CC RNA POLYMERASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; U12020; AAA75361.1; -.
DR EMBL; AF016461; AAC68834.1; -.
DR EMBL; X90711; CAA62242.1; -.
DR PIR; I40327; I40327.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; Baf; 1.
KW Transcription regulation; Activator.
SQ SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;

Query Match      9.7%; Score 129.5; DB 1; Length 267;
Best Local Similarity 31.0%; Pred. No. 0.0011;
Matches 49; Conservative 22; Mismatches 68; Indels 19; Gaps 6;

QY 83 IBPQIVGPMKTKGLNPKYDNPKEVGADR--TVNAVAATHLYGNPLIVVDFGTATTCYI 139
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
83 LRAQPLAMGLRNG-----YKPNQGLADRWACWGVLPVPSVHPPLLVASFATTLDTI 138

QY 140 DENKQYMGGAAPGTTISTEALYSRAAKLPRIETTRPDNII GKNTVSAMQSGILFGYVGQ 199
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
139 GPDNVFPGLIILPGPMRGALAYGTAHLPLADGLVADYPI--DTHQAIASGIA---AAQ 193

QY 200 VEGIVKRMKWAQ-----KQDLKVIATGGLAPLIANESD 232
DB :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
194 AGAIVR--QWLGRORYGOAPEIYVAGGQWPEVRQEA 229

RESULT 3
GRSB BACBR
ID _GRSB BACBR STANDARD; PRT; 4451 AA.
AC P14688;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Gramicidin S synthetase II [Includes: ATP-dependent proline adenylyase
DE (proA) (proline activase); ATP-dependent valine adenylyase (ValA)
DE (valine activase); ATP-dependent ornithine adenylyase (OrnA) (ornithine
DE activase); ATP-dependent leucine adenylyase (LeuA) (leucine activase)].
GN GRSB OR GRS2.
OS Bacillus brevis (Brevibacillus brevis).
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
OX NCB1_TaxID=1393;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 9399 / IFO 15520 / JCM 8504 / NCTC 7096;
RX MEDLINE=92219998; PubMed=1560782;
RA Turgay K., Krause M., Marahiel M.A.;
RT "Four homologous domains in the primary structure of GrsB are related
RT to domains in a superfamily of adenylyate-forming enzymes."
RL Mol. Microbiol. 6:529-546(1992).
RN [2]
RP SEQUENCE OF 1-948 FROM N.A.
RC STRAIN=Nagano;
RX MEDLINE=920411751; PubMed=1939016;
RA Hori K., Yamamoto Y., Tokita K., Saito F., Kurotsu T., Kanda M.,
RA Okamura K., Furuyama J., Saito Y.;
RT "The nucleotide sequence for a proline-activating domain of

```


CC IGP, AICAR and glutamate. The hlsF subunit catalyzes the
 CC cyclization activity that produces IGP and AICAR from PRFAR using
 CC the ammonia provided by the hlsH subunit (By similarity).
 CC -!- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulos-1-
 CC ylamino)methylideneamino]-1-[(5-phosphoribosyl)imidazole-4-
 CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
 CC aminoimidazole-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
 CC -!- PATHWAY: Histidine biosynthesis; fifth step.
 CC -!- SUBUNIT: Heterodimer of hlsH and hlsF (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE004927; AAG08525.1; -;
 CC PIR; H83002; H83002.
 CC HAWAP; MF 01013; -; 1.
 CC InterPro; IPR003009; FMY enzyme.
 CC InterPro; IPR006062; His biosynth.
 CC Pfam; PF00977; His biosynth; 1.
 CC TIGRFAMs; TIGR00735; hlsF; 1.
 CC Histidine biosynthesis; Lysase; Complete proteome.
 CC ACT_SITE 12 12 POTENTIAL.
 CC ACT_SITE 131 131 POTENTIAL.
 CC SQ SEQUENCE 256 AA; 27131 MW; 6486A76CD308FAA7 CRC64;

Query Match 6.6%; Score 88; DB 1; Length 256;
 Best Local Similarity 22.7%; Pred. No. 3;
 Matches 60; Conservative 45; Mismatches 79; Indels 80; Gaps 15;

Qy 5 IDVGNNTVLGVYHDKLEYHWRISRKTEDEFGMLRSLFDHSGLMFQIDGIIIS 64
 Db 11 LDVDNGRVVKGKFNIRDAGDPVEIARR-----YDEQGA--DEITFLDITA 55
 Qy 65 VV---PPIMFALERMCTKYFHIEPQIVGPMKTLGINKYDNPKEVGADRIVNAVAHLY 121
 Db .56 SVDGRDTHLTHTVERMASQVF--IPLTVGGVRSQDIR--NLLNAGADKVSINTAAVF-- 109
 Qy .122 GNPLIVD-----FGTATTCYIDENKQYMGAIAPGITISTEALYSRAAKLPRIE--- 173
 Db 110 -NPEFVGEADRFGSQCIIVDAIKK-----VSAPG-----EAPRWEIFTHG 150
 Qy 174 ---TRPDNII-----GKNTVSAM-QSGILFGY-VGQVGEIVKRMKQAKQDLKV 217
 Db 151 GRKPTGLDAVLWAKMEDLGAGEILLTSMDDQGVKSGVDLGVTRAISEAV-----NVPV 204
 Qy 218 IATGGL-----APLIANESDCI 234
 Db 205 IASGGVGNLEHLAAGILEGKADAV 228

RESULT 7

SAHH_METKA ID SAHH_METKA STANDARD; PRT; 424 AA.
 AC P58855;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Adenosylhomocysteinease (EC 3.3.1.1) (S-adenosyl-L-homocysteine
 DE hydrolase) (AdoHcyase).
 GN AHY OR MK0368.
 OS Methanopyrus kandleri.
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
 CC Methanopyrus.
 OK NCBI_TaxID=2320;
 RN [1]

RP SEQUENCE FROM N.A.
 RX STRAIN=AV19 / DSM 6324 / JCM 9639;
 RC MEDLINE=21927647; PubMed=11930014;
 RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
 RA Sherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
 RA Natsale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
 RA Malykh A.G., Koonin E.V., Kozayavkin S.A.;
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
 RT and monophyly of archaeal methanogens";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine
 CC + L-homocysteine.
 CC -!- COFACTOR: NAD (By similarity).
 CC -!- PATHWAY: Activated methyl cycle.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the adenosylhomocysteinease family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE010334; AAM01583.1; -;
 CC HAWAP; MF 00563; -; 1.
 CC InterPro; IPR000043; Ado_hcyase.
 CC Pfam; PF05221; AdoHcyase; 1.
 CC Pfam; PF00670; AdoHcyase; 1.
 CC TIGRFAMs; TIGR00936; ahcy; 1.
 CC PROSITE; PS00738; ADOHCYASE_1; FALSE_NEG.
 CC PROSITE; PS00739; ADOHCYASE_2; 1.
 CC Hydrolase; NAD; One-carbon metabolism; Complete proteome.
 CC KW Hydrolase; NAD; 213 244
 CC FT NP_BIND 213 244
 CC SQ SEQUENCE 424 AA; 47015 MW; 3DE7B10A3B44E1D5 CRC64;

Query Match 6.6%; Score 87.5; DB 1; Length 424;
 Best Local Similarity 20.8%; Pred. No. 5.8;
 Matches 62; Conservative 47; Mismatches 96; Indels 93; Gaps 14;

Qy 19 DGKLEYHWRISRKTEDEFGMLRSLFDHSGLMFQIDGIIISVVPPIMF-----AL 73
 Db 97 EGVHVVAWRGE-----TEBEY-----YQNDRLVL--SHEPDIIVDGDGDCI 135
 Qy 74 ERMCTKYFHIEPQIVGPMK--TGLN-----IKY--DNPKEVG 107
 Db 136 ARVHTEPDLAERVIGATEETTTGNRLHAMHREGVLKFPVIAVNDATKYLMDNRYGTG 195
 Qy 108 ADRIVNAVA--IHLNGNLIWVDFGTATTCYIDENKQYMG-GAIAPGITI---STE 160
 Db 196 QSALDGLMRATNILLACKTVVVVGYG---WCGRGIARRAGLGANVIVVEVDPIKAMEA 251
 Qy 161 LY-----SRAAKLPRIEITRPN--IIGKNTVSAMQSGILFGYVQVVE----- 201
 Db 252 IFDGRVWPMDDRAAEEGDIPITATGNRDVIRGEHIEKMKDGVILANAGHFDEIDKEYLE 311
 Qy 202 -----GIVKRMKQAKQDLKVIAATGGLAPLIANESDCIDIVDPFLTLKGLEL 248
 Db 312 EHCEEKIDRRGGLVTEYRMPDGKRVYLIAGRLVNLAAEGHPHIEIMDISFALQALSV 369

RESULT 8

PYRD_SALTI ID PYRD_SALTI STANDARD; PRT; 336 AA.
 AC Q827S9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Dihydroorotate dehydrogenase (EC 1.3.3.1) (Dihydroorotate oxidase)
 DE (DHODase) (DHODase) (DHOD).
 GN PYRD OR STY1079 OR T1862.
 OS Salmonella typhi.

KW Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FMN;
KT Complete proteome.
KW NP_BIND 290 FWN (POTENTIAL).
SQ SEQUENCE 336 AA; 36739 MW; 783F59D28DEE6398 CRC64;

Query Match
Best Local Similarity 21.6%; Score 87; DB 1; Length 336;
Matches 58; Conservative 36; Mismatches 79; Indels 96; Gaps 13;

QY 37 DDFGMLRSLRFHSGI--MFEQI-----DGIISSVPPIMPALERMCTKYFHIEPIQVG 89
DB 105 DAELGNRGFNGLVDNLVNENKAFHGDI- - - - - 136

QY 90 PGMKTGLNI--KYDNPKGVADRI VNAVAIAHLNGPLIV - - - - -VPFGTTAT 135
DB 137 -----GINIGKNKDTPVENGGDDYLCEKVAYAGYAIAINISSPTPLRLTLQYGDA-- 189

QY 136 YCYIDE-----NKQ-----YMGNAIPCIITISTEALSRAAKLPRIETRPDIII 180
DB 190 ----LDDLTAIRKKQNQLQAIIHKHYVPVAVKIAPDLCEEELIQVADSLLRN----DGI 243

QY 181 GKNTV-----SAMOSGILFGVVQGVEG--IVKRMMQOAKDLKVIATGGLAFLIA 228
DB 244 ATNYTLDRSLVOGMKCQCQTGLSGRPLQSKTEIIRLSOLKGQLPIGVGGIDSVIA 303

QY 229 NE-----SDCIDVIDPFUTLKXGLELIYE 251
DB 304 AREKAAGATLVQIYSGFI-FKGPLLIKE 331

RESULT 10
Y004.METAJ
ID Y004 METAJA STANDARD; PRT; 243 AA.

AC Q6031S;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DG Hypothetical protein MJ0004.
DN MJ0004.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcaceae;
OC Methanocaldococcaceae; Methanocaldococcus.
NCBI_TaxID=2190;
RX [1].

RN SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RE MEDLINE=96337999; PubMed=868087;
RA Sulton C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek K., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.P., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadov P.W., Hanna M.C.,
RA Cotton M.D., Roberts J.K., Hurd M.A., Kaine B.P., Borodovsky M.,
RA Kleen H.-P., Fraser C.M., Smith H.O., Woese C.R., Ventner J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RL Science 273:1058-1073(1996).

CC -! SIMILARITY: A.FERMENTANS ACTIVATOR OF (R)-HYDROXYGLUTARYL-COA
DEHYDRATASE (HGDC), TO E.COLI YJIL AND M.JANNASCHII MJ0800.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to licens@isb-sib.ch).
CC -----
CC EMBL; U67459; AB97985.1; --
CC PIR; D64300; D64300.
CC KASP; PI1568; LHUX.
CC TIGR; MJ0004; --

DR InterPro; IPR002731; ATPase_BadF.
DR Pfam; PF01869; BCrad_BadFg_1.
DR ProDom; PD006344; ATPase_BadF_1.
DR TIGRFams; TIGR00241; Coa_E_activ_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 243 AA; 26897 MW; 288C4F4190575E53 CRC64;

Query Match
Best Local Similarity 21.3%; Score 86.5; DB 1; Length 243;
Matches 54; Conservative 44; Mismatches 81; Indels 75; Gaps 12

QY 1 LLIVDVGTNVTVLGYYHDCKLEYHWRIETS SHKHTEDEFEGMILRSLEFDHSGLMFPEIDI 60
DB 1 MILGIDVGSTTTKMVLMSDKIIW-----YKIED-----IGV 32

QY 61 IISSVPPPMPFALERMCTKYEHIEPOIVGPQMKTGLNIKYPNPKEVGADRIVNAAAI-- 118
DB 33 VBEE---DILLNMVKIEIOKPID-KIVATG-----YGRHKVSFPADKIVEPVIALGK 80

QY 119 ---HLXGNPLIIVVDF-GTATTTCYIDENKQVMGAIPA GITITSEALYSRA---AKLPRI 171
DB 81 GANYFFNEADVIGDIGGDQTKVLKIDKNGKVDFILSDKCAAGTKFLEKALKLDIKDN 140

QY 172 EIT-RPDINIIGNVTSM-----OSGILFG-----YVGQEVGIVKRMKWQ 210
DB 141 EINRYKSNDIAKITSSMCVAFABESIIISLLSKKPKREGILMVYESIINRVIMTNLRLIK 200

QY 211 AKQDLKVIATGGLA 224
DB 201 -----NIVPSGGVA 209

RESULT 11
PYRD.ECOL6
ID PYRD.ECOL6 STANDARD; PRT; 336 AA.

AC Q8FJ91;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Dihydrooorotate dehydrogenase (EC 1.3.3.1) (dihydrooroate oxidase)
DE (DHODase) (DHODase) (DHOD).
GN PYRD OR C1081.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RX [1].

RN SEQUENCE FROM N.A.
RC STRAIN=06:H1 / CF7073 / ATCC 700928;
RE MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perla N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
CC -! CATALYTIC ACTIVITY: (S)-dihydrooorotote + O(2) = orotate +
H(2O)(2).
CC -----
CC -! COFACTORS: Binds 1 FMN per subunit (By similarity).
CC -! PATHWAY: Pyrroimidine biosynthesis; fourth step.
CC -! SUBUNIT: Homodimer (By similarity).
CC -! SUBCELLULAR LOCATION: Inner side of the membrane (By similarity).
CC -! SIMILARITY: Belongs to the dihydrooorotate dehydrogenase family.
Subfamily 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to licens@isb-sib.ch).
CC -----
CC EMBL; U67459; AB97985.1; --
CC PIR; D64300; D64300.
CC KASP; PI1568; LHUX.
CC TIGR; MJ0004; --

```

CC -----
DR EMBL; AE016758; AAN79549.1; ALT_INIT.
DR HAMAP; MF_00225; -, 1.
DR InterPro; IPR001295; DHO_dh.
DR InterPro; IPR005719; DHO_dh2.
DR InterPro; IPR003009; FMN_enzyme.
DR Pfam; PF01180; DHODEHase; 1.
DR TIGRFAMs; TIGR01036; pyrd_sub2; 1.
DR PROSITE; PS00911; DHODEHASE_1; 1.
DR PROSITE; PS00912; DHODEHASE_2; 1.
DR Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FMN;
KW Complete proteome.
FT NP_BIND 290 298 FMN (POTENTIAL).
SQ SEQUENCE 336 AA; 36817 MW; A8B92A71ED2EED5B CRC64;

Query Match 6.4%; Score 85; DB 1; Length 336;
Best Local Similarity 25.0%; Pred. No. 7.2;
Matches 51; Conservative 27; Mismatches 70; Indels 56; Gaps 11;

QY 95 GLNI--KYDNPKEVGADRIVNAVAATHLYGNPLIV-----VDGTTATTCYID 140
DB 137 GINIGKNDTPVEQKDDYLICHEKIYAYAGYAINISSNTPGLRTLQVGEA-----LD 191
QY 141 E-----NKK-----YMGGAIAFGITISTEALYSRAKLPRIETRPDNIIGKNT- 184
DB 192 DLLTAIKNKQNDLVQVHHKVPVIAVKIAPDLSEELIQVADSLVRHNI---DGVIAITNT 248
QY 185 -----VSAM-----QSGILFGVQVQEG--IVKMKQAKQDLKVIATGGLAPLIANE--- 230
DB 249 LDRSLVQGMKNCQDTGSLGRPLQLKSTRIIRLSQELNRLPIIGVGGIDSVAAREKI 308
QY 231 ---SDCIDIVDPFLTLKGLLEIYE 251
DB 309 AGASLVQIYSGFI-FKGPPLIKE 331

RESULT 12
PYRD_ECOLI STANDARD; PRT; 336 AA.
AC P05021;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Dihydroorotate dehydrogenase (EC 1.3.3.1) (Dihydroorotate oxidase)
DE (DHODEHase) (DHOD)
GN PYRD OR B0945 OR Z1294 OR ECS1029.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-7.
RX MEDLINE=85285014; PubMed=2992959;
RA Larsen N.J., Jensen K.F.;
RT "Nucleotide sequence of the pyrd gene of Escherichia coli and
RT characterization of the flavoprotein dihydroorotate dehydrogenase.";
RL Eur. J. Biochem. 151:59-65(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,

```

```

RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishio Y., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takai H., Honda T., Sasaki C., Ogawara N., Yasunaga T.,
RA Kuwara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -1- CATALYTIC ACTIVITY: (S)-dihydroorotate + O(2) = orotate +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 FMN per subunit.
CC -1- PATHWAY: Pyrimidine biosynthesis; fourth step.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Inner side of the membrane.
CC -1- SIMILARITY: Belongs to the dihydroorotate dehydrogenase family.
CC -1- Subfamily 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X02826; CAA26594.1; -
CC EMBL; AE000196; AAC74031.1; -
CC EMBL; D90732; BAA35700.1; -
CC EMBL; AE005284; AAG55431.1; -
CC EMBL; AF002553; BAB34452.1; -
CC PIR; A23109; DEECDO.
CC PIR; C85621; C85621.
CC PIR; E90757; E90757.
CC PDB; 1F76; 16-OCT-02.
CC SWISS-2DPAGE; P05021; COLI.
CC EcoGene; EGI0807; pyrd.
CC HAMAP; MF_00225; -, 1.
CC InterPro; IPR001295; DHO_dh.
CC InterPro; IPR005719; DHO_dh2.
CC InterPro; IPR003009; FMN_enzyme.
CC Pfam; PF01180; DHODEHase; 1.
CC TIGRFAMs; TIGR01036; pyrd_sub2; 1.
CC PROSITE; PS00911; DHODEHASE_1; 1.
CC PROSITE; PS00912; DHODEHASE_2; 1.
KW Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FMN;
KW Complete proteome; 3D-structure.
FT NP_BIND 290 298 FMN (POTENTIAL).
SQ SEQUENCE 336 AA; 36774 MW; 973227BAE6B83622 CRC64;

Query Match 6.4%; Score 85; DB 1; Length 336;

```



```

RP MUTAGENESIS.
RX MEDLINE=91355180; PubMed=1653010;
RA Rothery R.A., Weiner J.H.;
RT "Alteration of the iron-sulfur cluster composition of Echerichia
RT coli dimethyl sulfoxide reductase by site-directed mutagenesis.";
RL Biochemistry 30:8296-8305(1991).
CC -1- FUNCTION: Electron transfer subunit of the terminal reductase
CC during anaerobic growth on various sulfoxide and N-oxide
CC compounds.
CC -1- COFACTOR: Binds 4 4Fe-4S clusters.
CC -1- SUBUNIT: Heterotrimeric enzyme composed of a catalytic heterodimer
CC (DmeAB) and a membrane anchor protein (DmsC).
CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; J03412; AAA83844.1; -.
DR EMBL; AE000191; AAC73981.1; -.
DR EMBL; D90727; BAA35627.1; -.
DR PIR; F64828; F64828.
DR HSP; P00195; 1CLF.
DR EcoGene; EGI0233; dmsB.
DR InterPro; IPR001450; 4Fe4s_ferredoxin.
DR Pfam; PF00037; fer4; 1.
DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
KW Electron transport; 4Fe-4S; Iron-sulfur; Complete proteome.
FT INIT MET 0 0
FT METAL 13 13 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 16 16 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 19 19 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 23 23 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 66 66 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 69 69 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 74 74 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 78 78 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 98 98 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 101 101 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 104 104 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 108 108 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 125 125 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT METAL 128 128 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT METAL 140 140 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT METAL 144 144 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT METAL 101 101 C->F S.W.Y. LOSS OF ELECTRON TRANSFER
FT MUTAGEN FROM MENAQUINOL TO DMSO.
FT CONFLICT 169 169 P -> PRA (IN REF. 1).
FT SEQUENCE 204 AA; 22738 MW; 90425D307DEF2130 CRC64;
Query Match 6.3%; Score 83.5; DB 1; Length 204;
Best Local Similarity 23.3%; Pred.No.5.4;
Matches 45; Conservative 27; Mismatches 66; Indels 55; Gaps 10;
QY 15 GVYHDKLEYYHRIETSRHKHTEDFGCMILRSFLDHSGLMPEQIDGIIISVVPPIMFALE 74
DB 51 GVWHQNVFAY--LSISCNHEDP-----ACTKVCPSGMHKREDGFVVVD-----E 95
QY 75 RMC--TKYPHIEPQIVGPGMKTGLNLYKNPK-----EVGADRIVNAVAAILHYNPLI 126
DB 96 DVCIGCRYCHMACPYGAP-----QYNETKGHMTKDCGCDYDRAVEAGKKPVCVSCPLR 147
QY 127 VVDFGPTATTCYIDENKQYWG--GATARGITISTEALYSRAAKLPRIETRPDNIIGKNT 184
DB 148 ALDFFGP-----IDELRKHGDLAAVAP-----LPRAHFTKP-NIVIKPN 185
QY 185 VVAQSGILFGVV 197

```

